New Mechanisms of Activation by Histone Demethylases in Gene Regulation

New Mechanisms of Activation by Histone Demethylases in Gene Regulation

A dissertation presented

by

Erin Amelia Clark

To

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological Chemistry and Molecular Pharmacology

Harvard University

Cambridge, Massachusetts

July 2013

Dissertation Advisor: Yujiang Geno Shi Erin Amelia Clark

New Mechanisms of Activation by Histone Demethylases in Gene Regulation

Abstract

The epigenetic mechanisms that connect hormone signaling to chromatin remain largely unknown. Here we show that LSD1/KDM1A is a critical glucocorticoid receptor (GR) coactivator and report a previously unexplored mechanism where LSD1 activates gene transcription through H3K4me2 demethylation. We demonstrate that direct interaction of GR with LSD1 primarily inhibit its activity against H3K4me1 *in vitro*. While this interaction enables GR to recruit LSD1 *in vivo* and allows loss of H3K4me2, it impedes further demethylation. Thus resulting in conversion of H3K4me2 to H3K4me1 at enhancers and promotes H3K27 acetylation and gene activation. We also find that H3K4me2 is an early enhancer mark predicting GR and LSD1 recruitment. These findings differ from the reported mechanism for ER and AR-mediated gene activation, providing a novel mechanism for LSD1 coactivator function as well as shed light on the role of H3K4me2 and enhancers in hormone-mediated gene regulation. In addition we present evidence supporting never before characterized H3K79me3 demethylase activity by members of the JMJD2 family of proteins.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I am incredibly lucky to have many people to acknowledge for their help, support, and contributions, which have made this work possible. I want to first thank my advisor Dr. Yujiang (Geno) Shi, for teaching me many important lessons on what it means to be a PI and an advisor as well as for giving me the tools to carry out this research. Many thanks go to several past Shi Lab members. Two in particular are Rui Fang, who provided much needed as well as very wise guidance and support; and to Paco Kang, who worked closely with me on several experiments and made the lab a more enjoyable work environment through lively discussions on myriad topics as well as the occasional pyrotechnic display to keep me on my toes.

Outside of the lab are many other wonderful people who have been tremendously helpful in innumerable ways. I have had the honor and pleasure of being an active teacher in the Harvard community working with and learning from some amazing people, namely Cheryl Vaughn, Sarah Wojiski, and Monideepa Roy. David Cardozo and his team of Lisa Rossini, Bob Bridges and Chelsea Noriega deserves a special thank you for their tireless support of DMS students as they endeavor to succeed in many types of careers following graduate school. Thank you also to Johanna Kowalko, Tom Ennis and Allison Lau for their friendship and collaboration as fellow Education Path Directors. These folks have all been a major influence on the development of my career trajectory as well as my development as a human being.

Many thanks also go to my dissertation advisory committee, Bob Kingston, Anders Näär, and Raul Mostoslavsky for their instrumental advice and support. Thank you also to my defense committee, Danaesh Moazed, Kami Ahmad, and Laurie Boyer for their thoughtful questions and comments.

Last but not least I want to thank my friends and family. I would not have made it all the way to a PhD program at Harvard Medical School, let alone through the program without the love and endless support of my mother and father Carin and James Clark as well as my sister Lauren Clark. A special thank you to my wonderful friends Scott Jones, Ben Wolfe, Christina Agapakis, Nick Seaver, Hilary and Matt Eaton who have made the bad times a little less bad and the good times all that much better.

Chapter 1

Introducing Epigenetics and Histone Methylation

The Far Reaching Umbrella of Epigenetics

Epigenetics is a newly emerged field with wide ranging impact on our understanding of biology. Epigenetics is generally accepted as the study of changes in gene expression or cellular phenotype that are not the result of changes in DNA sequence, some of which are heritable. The Greek root epi- implies mechanisms that occur on top of the genetic code, which also implies shared characteristics with classical genetics such as heritability. However, the origin of the word epigenetics has its true roots in an attempt to describe the connections between genetics and development in the late 1930's by relating it to the word epigenesis [1]. This lead to observations of a "new type of inheritance" occurring across mitotic cell divisions [2,3] and later across generations [4] being intermingled with studies of genetic "switches" responsible for turning genes on/off during development. These two lines of research have come together to form the modern epigenetic amalgamation, which encompasses processes as diverse as DNA methylation to nuclear architecture.

While there is good evidence supporting the existence of new non-genetic mechanisms of heritability, the mechanism(s) that transfer epigenetic information are not understood except in a few cases, such as for DNA methylation. On the other hand, our understanding of the molecular and biochemical switches that flip genes on and off has expanded dramatically since 1990's. We now understand that myriad mechanisms, both nuclear and non-nuclear, affect the way genes are expressed, and most of these mechanisms have been

included under the umbrella of epigenetics. This helps explain the ambiguous definition of epigenetics as well as the debate in some circles about over whether inheritance is a requirement for the definition. While it remains to be determined whether heritable versus non-heritable changes in gene expression will eventually subdivide the field, one thing is certain; the epigenetic umbrella has many spokes all of which have proven biologically and clinically important.

Similar to discoveries of genetic mutations underlying disease, alteration of the epigenome has been similarly implicated in a wide variety of human diseases such as cancer, cognitive dysfunction, reproductive disorders, respiratory, autoimmune, and neurobehavioral diseases, as well as contributing to the aging process and evolution [5-12]. Because epigenetics allows environmental signals to alter the function of the genome it is also thought to contribute heavily to diseases such as heart disease, obesity and diabetes, which we now understand to be complex mixtures of environmental/lifestyle exposures and genetic susceptibilities.

Epigenetic mechanisms allow the environment outside both the cell and the organism to influence gene expression. Environmental stimuli are detected and processed by an organism. In the case of multicellular organisms this signal is then transmitted to pertinent tissues and cell-types which respond in part by transmitting that signal to the nucleus where regulatory factors alter the pattern of gene expression. This allows an organism to respond to its environment down to the individual cell. An excellent example of this process is the human body's response to stress.

When a human encounters stress, the adrenal cortex releases hormones, one of which is cortisol. Cortisol is part of the glucocorticoid family of steroid hormones and circulates throughout the body to affect multiple organ systems. These effects include depressing the immune system and inflammation as well as mobilizing glucose. Cortisol carries out these effects in large part through binding to nuclear hormone receptors located in the cytoplasm of cells such as immune and liver cells. For example the glucocorticoid receptor (GR) is sequestered in the cytoplasm of many cell-types and upon binding to a ligand such as cortisol will translocate to the nucleus where it acts as a hormone-dependent transcription factor regulating specific gene targets [13]. mRNA profiling of immune cells and hepatocytes responding to glucocorticoids reveals activation and suppression of genes involved in inflammatory signaling, glucose metabolism, and cell survival [14,15]. These changes in gene expression of various cell and tissue types produce a coordinated response by the organism to the instigating stressor.

These processes are also critical during development where patterns of cell and tissue differentiation are directed by diffusible signaling molecules. Studies characterizing the epigenome throughout development have demonstrated significant alteration of the epigenetic landscape as cells specialize and thus restrict gene expression to a specific subset of genes [16,17]. At various stages newly established epigenetic landscapes are maintained for many cellular generations and in some cases throughout the lifetime of the organism generating the foundation for cell and tissue differentiation.

However, unlike the long-term changes in the epigenome during development, similar mechanisms are also used on shorter time-scales. Returning to the example of stress, epigenetic switches temporarily alter gene expression allowing for dynamic adaptation to the environment. Most often, removal of the hormone signal will reset the switches allowing cells to revert to baseline function. However, our understanding of the long-lasting effects of chronic exposure to altered physiological states and the contributions of stabile epigenetic changes are beginning to be explored [18,19].

Chromatin

Many epigenetic mechanisms center on the structure and function of the protein DNA complex known as chromatin. Even in its most unadorned state, the eukaryotic genome in nature is a complex structure of histone proteins wrapped in DNA, a single unit of which is called a nucleosome (Figure 1.1). A nucleosome consists of an octamer of histone proteins (two of each H3, H4, H2A, H2B) wrapped in approximately 147 bp of DNA [20-23]. Chromatin is a dynamic structure with two basic states known as euchromatin and heterochromatin. The two states are defined relative to each other with euchromatin being more loosely packaged allowing access to the DNA template, and heterochromatin being a more condensed structure with nucleosomes tightly packed together restricting access. The most dramatic demonstration of these two states occurs during the cell cycle where the genome goes from a mixture of tightly and loosely packaged regions during interphase to the most extremely condensed state visible as individual chromosomes during mitosis.

At the level of the nucleosome there are several biochemical processes included under the umbrella of epigenetics. Firstly, DNA itself can be chemically modified. In mammals the addition of a methyl group to the 5th position of cytosine was one of the first epigenetic marks to be discovered [24,25]. Despite more than 65 years studying DNA methylation and the enzymes responsible for adding this modification, it is only in the last five years that we are beginning to understand how this mark is dynamically regulated. The recent discovery of DNA hydroxymethylases and subsequent modifications has shed new light on an old assumption about the long-term stability of DNA methylation [26]. Increasing evidence suggests DNA methylation, while perhaps not as dynamic as other chromatin modifications, is in fact reversible and actively regulated with significant implications for physiology and disease (reviewed in [27]).

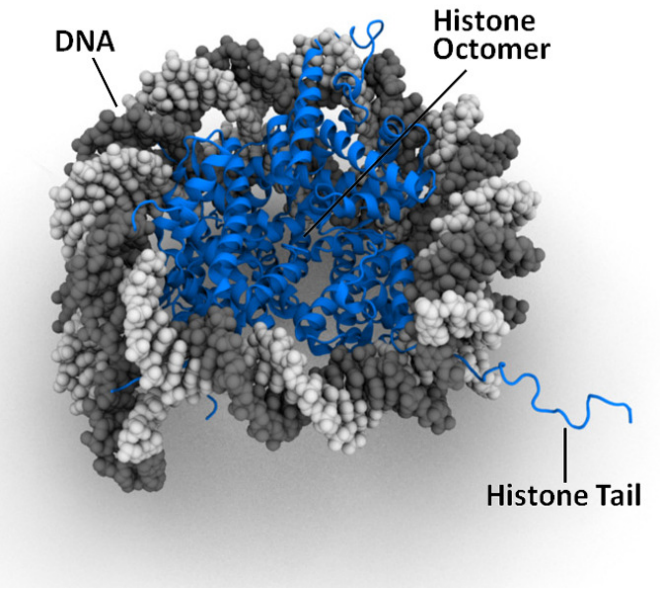


Figure 1.1 Basic 3D structure of a single nucleosome. DNA is represented by gray space-filling model and histone proteins are represented by blue ribbon models. Histone tails can be seen protruding from histone-DNA complex.

(Image modified from James Hedberg original licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.)

Up one level from DNA we find myriad post-translational modifications to histone proteins. The majority of these modifications occur on the N-terminal tail, which is an unstructured region of the histone protein protruding from the overall nucleosome structure (Figure 1.1). Many covalent modifications to histone tails have been described (acetylation, methylation, phosphorylation, sumolation, and ubiquitination) and tails can be modified at multiple amino acid residues allowing for numerous and diverse combinations. Histone tails decorated with various marks signal information to the cell about the chromatin and transcriptional state of that region of the genome. Studies have also established that altering these covalent histone modifications directly and/or indirectly influences chromatin through the recruitment of effector proteins to specific chromatin regions [28-30]. Similar to a system of color-coded tabs in your favorite book, histone modifications form a type of code on top of the genetic code indicating which regions and genes should be freely accessible and expressed and which regions should be silenced (as well as all manner of intermediate states).

While we have only begun to scratch the surface in understanding the histone code, some general trends have emerged. For example, histone acetylation is generally associated with euchromatin and gene activation. This in large part due to the additional negative charge added to the histone with this chemical modification (Figure 1.2 A). The increased negative charge helps repel the also negatively charged DNA opening chromatin up. However, other modifications such as methylation are less clear-cut. Methylation itself is neutral and therefore does not affect the biochemical properties of histone DNA interactions (Figure 1.2 B). Instead the impact of histone methylation on chromatin depends entirely on the

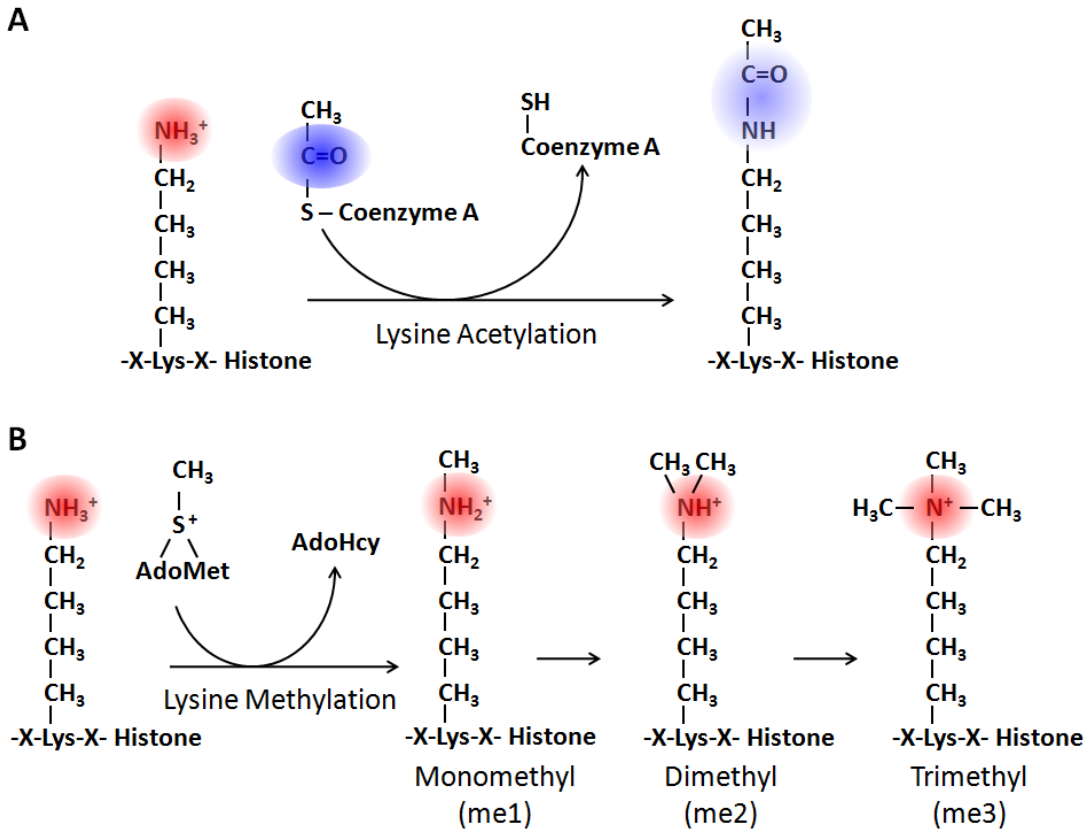


Figure 1.2 Lysine modifications.

(A) Histone lysine residues can be acetylated which changes the overall charge of the histone protein. (B) Histone lysine residues can also be methylated to three different degrees which does not change the charge of the protein.

type of proteins that recognize this mark. Proteins that recognize specific modifications are called “readers” of the histone code. In terms of gene expression, the function of a histone reader depends on whether it possesses an enzymatic activity and/or recruits other proteins that favor open or closed chromatin. Decoding histone methylation has proven to be a daunting task. Methylation occurs at many residues (Figure 1.3), including one lysine in the globular region of the histone (H3K79), and to varying degrees with the addition of

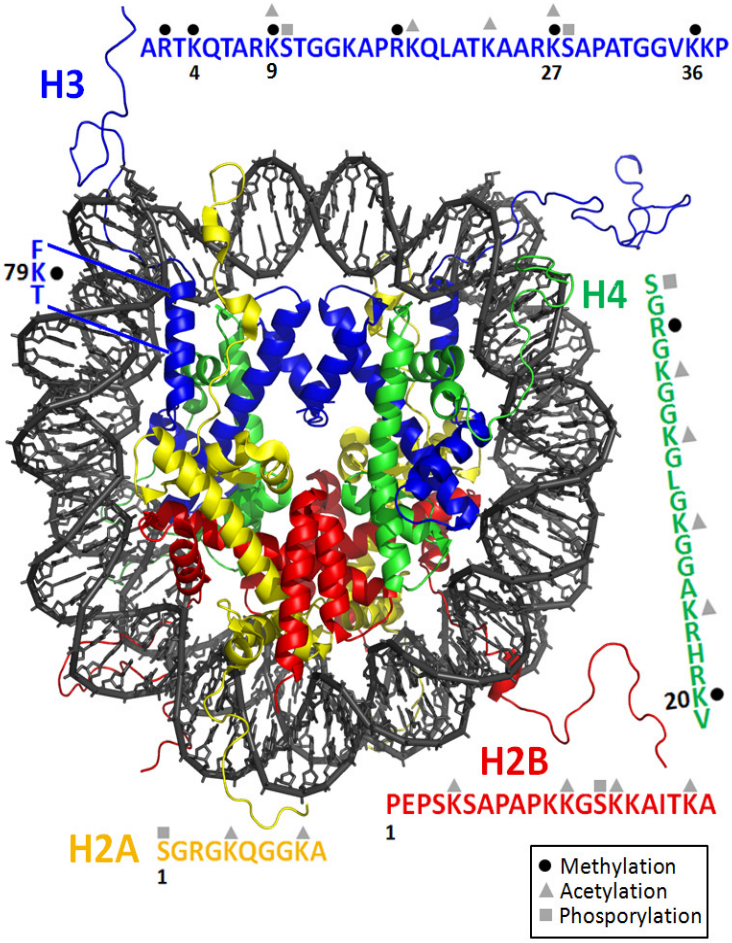


Figure 1.3 Histone tail modifications.

Crystal structure of a nucleosome with histones H2A shown in yellow, H2B in red, H3 in blue, and H4 in green. Tail residues are spelled out with known sites of methylation, acetylation and phosphorylation indicated.

one, two or three methyl groups to a single residue (mono, di, or trimethylation) (Figure 1.2 B). Histone methylation has been implicated in euchromatin and heterochromatin formation as well as many stages of transcriptional regulation from initiation to mRNA processing all dependent on the residue and degree of methylation. Table 1.1 summarizes

the function of known methylation sites, but certainly represents an incomplete understanding of this modification.

Table 1.1 Histone Methylation

Histone	Methylation Site	Function
H1	K26	Transcription silencing
H3	K4	Transcription activation
	R8	Transcription silencing
	K9	Transcription silencing or activation, imprinting, DNA methylation
	R17	Transcription activation
	K27	Transcription silencing, X-inactivation
	K36	Transcription activation (elongation)
	K79	Transcription activation (elongation), DNA repair/checkpoint response
H4	R3	Transcription activation
	K20	Transcription silencing or activation, checkpoint response
	K59	Transcription silencing

Just as there are readers of the histone code there are also "writers" and "erasers" of the code. These are enzymes responsible for regulating when and where histone modifications are laid down. The balance between writer and eraser activities dynamically regulates the histone code throughout processes such as development and cellular responses to stimuli. The writers and erasers for histone methylation are two classes of enzymes known as histone methyltransferases (HMTs, writers of histone methylation) and histone demethylases (HDMs, erasers of histone methylation). While the mechanisms of histone acetylation and deacetylation have been well studied for many years, the first HMT wasn't

discovered until fairly recently [31] and up until the discovery of the first HDM (LSD1/KDM1a) three years later, histone methylation was thought to be an irreversible modification [32].

Following the discovery of LSD1, many more histone demethylases have been characterized solidifying histone methylation as a dynamic regulatory mechanism involving a balance between numerous writers and erasers. There are two major classes of histone demethylases. The first, so far only includes two proteins, LSD1 and LSD2, which carry out demethylation via a flavin dependent monamine oxidase reaction. The second class mediates oxidative demethylation by radical attack catalyzed by a Jumonji C (JMJC) domain [33]. While a large number of histone methyltransferases have been described, fewer of their counterpart demethylases have been identified. Until very recently no enzyme had been shown to demethylate arginine residues of histone tails [34], and some, such as methylation on histone H3 lysine 79 and H4 lysine 20 (H3K79 and H4K20), lack a characterized demethylases all together. This disparity leads many to believe there are novel histone demethylases left undiscovered along with no doubt many facets of the epigenetic code [35]. Table 1.2 summarizes the known HDMs. Together DNA, histone, and nucleosome modification form an intricate and vastly unexplored network of interacting mechanisms that regulate chromatin structure and gene expression, both of which contribute significantly to the biology of major fields of research such as development and stem cells, as well as numerous human diseases and disorders [36].

Table 1.2 Histone demethylases

HDM	Synonyms	Substrates
LSD histone demethylases		
LSD1	AOF2, BHC110, KDM1A	H3K4me1/2, H3K9me1/2
LSD2	AOF1, KDM1B	H3K4me1/2
JMJC histone demethylases		
JMJD5	KDM8	H3K36me2
JMJD6	PSR, PTDSR	H3R2, H4R3
FBXL10	JHDM1B, KDM2B	H3K36me1/2/3
FBXL11	JHDM1A, KDM2A	H3K36me1/2
KIAA1718	JHDM1D	H3K9me1/2, H3K27me1/2
PHF8	JHDM1F	H3K9me1/2, H4K20me1
PHF2	JHDM1E	H3K9me2
JMJD1A	JHDM2A, TSGA, KDM3A	H3K9me1/2
JMJD3	KDM6B	H3K27me2/3
UTX	KDM6A	H3K27me2/3
JMJD2A	JHDM3A, KDM4A	H3K9me2/3, H3K36me2/3, H1.4K26me2/3
JMJD2B	JHDM3B, KDM4B	H3K9me2/3, H3K36me2/3, H1.4K26me2/3
JMJD2C	JHDM3C, GASC1, KDM4C	H3K9me2/3, H3K36me2/3, H1.4K26me2/3
JMJD2D	JHDM3D, KDM4D	H3K9me2/3, H3K36me2/3, H1.4K26me2/3
JARID1A	RBP2, KDM5A	H3K4me2/3
JARID1B	PLU1, KDM5B	H3K4me2/3
JARID1C	SMCX, KDM5C	H3K4me2/3
JARID1D	SMCY, KDM5D	H3K4me2/3
NO66		H3K4me2/3, H3K36me2/3

One appealing aspect of epigenetic research is the malleability of the epigenome through manipulation of protein and enzyme regulators. The advent of human genetics and advances in virology and recombinant DNA technologies has made the possibility of correcting genetic mutations that underlie human disease a plausible and auspicious goal. Unfortunately, despite being first proposed over 40 years ago, gene therapy has been slow to break into the clinical setting as a safe and effective treatment. However, the paradigm shift in treatment strategy associated with gene therapy does not impede the development of epigenetic therapies given that enzymes are medical researchers' favorite targets. Thus epigenetic therapies combine drugable targets with altering gene expression which have as significant an impact on physiology as direct mutation. In addition, the potential for long-lasting transmissibility of epimutations makes the epigenetic code an incredibly promising area of biomedical research.

References

1. Waddington, C.H. An introduction to modern genetics, 2 p.l., [7]-441 p. (The Macmillan company, New York,, 1939).

2. Riggs, A.D. X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet 14, 9-25 (1975).

3. Holliday, R. & Pugh, J.E. DNA modification mechanisms and gene activity during development. Science 187, 226-32 (1975).

4. Kaati, G., Bygren, L.O. & Edvinsson, S. Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. Eur J Hum Genet 10, 682-8 (2002).

5. Bilinski, P. et al. Epigenetic regulation in drug addiction. Ann Agric Environ Med 19, 491-6.

6. Cosentino, C. & Mostoslavsky, R. Metabolism, longevity and epigenetics. Cell Mol Life Sci 70, 1525-41.

7. Dang, M.N., Buzzetti, R. & Pozzilli, P. Epigenetics in autoimmune diseases with focus on type 1 diabetes. Diabetes Metab Res Rev 29, 8-18.

8. Diaz-Garcia, C., Estella, C., Perales-Puchalt, A. & Simon, C. Reproductive medicine and inheritance of infertility by offspring: the role of fetal programming. Fertil Steril 96, 536-45.

9. Gomez-Diaz, E., Jorda, M., Peinado, M.A. & Rivero, A. Epigenetics of host-pathogen interactions: the road ahead and the road behind. PLoS Pathog 8, e1003007.

10. Lagali, P.S., Corcoran, C.P. & Picketts, D.J. Hippocampus development and function: role of epigenetic factors and implications for cognitive disease. Clin Genet 78, 321-33.

11. Roth, T.L., Roth, E.D. & Sweatt, J.D. Epigenetic regulation of genes in learning and memory. Essays Biochem 48, 263-74.

12. Suva, M.L., Riggi, N. & Bernstein, B.E. Epigenetic reprogramming in cancer. Science 339, 1567-70.

13. Yamamoto, K.R. Steroid receptor regulated transcription of specific genes and gene networks. Annu Rev Genet 19, 209-52 (1985).

14. Phuc Le, P. et al. Glucocorticoid receptor-dependent gene regulatory networks. PLoS Genet 1, e16 (2005).

15. Galon, J. et al. Gene profiling reveals unknown enhancing and suppressive actions of glucocorticoids on immune cells. FASEB J 16, 61-71 (2002).

16. Xie, W. et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell 153, 1134-48.

17. Serrano, L., Vazquez, B.N. & Tischfield, J. Chromatin structure, pluripotency and differentiation. Exp Biol Med (Maywood) 238, 259-70.

18. Zovkic, I.B. & Sweatt, J.D. Epigenetic mechanisms in learned fear: implications for PTSD. Neuropsychopharmacology 38, 77-93.

19. Heinzelmann, M. & Gill, J. Epigenetic Mechanisms Shape the Biological Response to Trauma and Risk for PTSD: A Critical Review. Nurs Res Pract 2013, 417010.

20. Van Holde, K.E., Allen, J.R., Tatchell, K., Weischet, W.O. & Lohr, D. DNA-histone interactions in nucleosomes. Biophys J 32, 271-82 (1980).

21. Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. & Richmond, T.J. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389, 251-60 (1997).

22. Kornberg, R.D. & Lorch, Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell 98, 285-94 (1999).

23. Zhang, K. & Dent, S.Y. Histone modifying enzymes and cancer: going beyond histones. J Cell Biochem 96, 1137-48 (2005).

24. Vischer, E., Zamenhof, S. & Chargaff, E. Microbial nucleic acids; the desoxypentose nucleic acids of avian tubercle bacilli and yeast. J Biol Chem 177, 429-38 (1949).

25. Griffith, J.S. & Mahler, H.R. DNA ticketing theory of memory. Nature 223, 580-2 (1969).

26. Tahiliani, M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324, 930-5 (2009).

27. Franchini, D.M., Schmitz, K.M. & Petersen-Mahrt, S.K. 5-Methylcytosine DNA demethylation: more than losing a methyl group. Annu Rev Genet 46, 419-41.

28. Kouzarides, T. Chromatin modifications and their function. Cell 128, 693-705 (2007).

29. Martin, C. & Zhang, Y. The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol 6, 838-49 (2005).

30. Strahl, B.D. & Allis, C.D. The language of covalent histone modifications. Nature 403, 41-5 (2000).

31. Rea, S. et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature 406, 593-9 (2000).

32. Shi, Y. et al. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell 119, 941-53 (2004).

33. Tsukada, Y. et al. Histone demethylation by a family of JmjC domain-containing proteins. Nature 439, 811-6 (2006).

34. Chang, B., Chen, Y., Zhao, Y. & Bruick, R.K. JMJD6 is a histone arginine demethylase. Science 318, 444-7 (2007).

35. Shi, Y. & Whetstine, J.R. Dynamic regulation of histone lysine methylation by demethylases. Mol Cell 25, 1-14 (2007).

36. Papp, B. & Plath, K. Epigenetics of reprogramming to induced pluripotency. Cell 152, 1324-43.

Chapter 2

H3K4me2 Demethylation by LSD1/KDM1A in Glucocorticoid Target Gene Activation

Erin A. Clark[1], Wenqi Xu[2], Feizhen Wu[2], Paco Kang[1], Rui Fang[1], Weiguo Zou[3], Yujiang Geno Shi[1]

1. Division of Endocrinology, Diabetes and Hypertension, Departments of Medicine and BCMP, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA
2. Laboratory of Epigenetics, Institutes of Biomedical Science, Fudan University, Shanghai 200032, P.R. China
3. Department of Immunology and Infectious Diseases, Harvard School of Public Health, Department of Medicine, Harvard Medical School, and Ragon Institute of MGH, Harvard and MIT, Boston, Massachusetts, USA.

Author Contributions

The work presented in this chapter was conceptualized by Dr. Yujiang Shi and Erin Clark. The biochemical, molecular, and cellular experimentation was primarily conducted by Erin Clark. Wenqi Xu and Dr. Feizhen Wu preformed all the deep sequencing and analysis of was carried out in conjunction with Erin Clark. Paco Kang contributed to the cloning of truncated proteins used in biochemical characterizations. Dr. Rui Fang provided technical support for most of the methods employed. Dr. Weiguo Zou carried out the culturing and differentiation experiments of mouse bone marrow stromal cells isolated.

Abstract

The epigenetic mechanisms that connect hormone signaling to chromatin remain largely

unknown. Here we show that LSD1/KDM1A is a critical glucocorticoid receptor (GR) coactivator and report a previously unexplored mechanism where LSD1 activates gene transcription through H3K4me2 demethylation. We demonstrate that direct interaction of GR with LSD1 primarily inhibit its activity against H3K4me1 *in vitro*. While this interaction enables GR to recruit LSD1 *in vivo* and allows loss of H3K4me2, it impedes further demethylation. Thus resulting in conversion of H3K4me2 to H3K4me1 at enhancers and promotes H3K27 acetylation and gene activation. We also find that H3K4me2 is an early enhancer mark predicting GR and LSD1 recruitment. These findings differ from the reported mechanism for ER and AR-mediated gene activation, providing a novel mechanism for LSD1 coactivator function as well as shed light on the role of H3K4me2 and enhancers in hormone-mediated gene regulation.

Background

Glucocorticoids (GCs) are essential hormones (aka stress hormones) released from the adrenal cortex that affect many physiological processes. Cortisol, the natural GC in humans, is known to influence glucose and lipid metabolism, bone homeostasis, stress response, development, the immune system, as well as behavior. At the cellular level, GCs impact many processes including proliferation, apoptosis, and differentiation. Clinically, GC imbalance has been implicated in a wide-range of diseases while GC restraint on the immune system makes them some of the most potent treatments for inflammatory and immune diseases [1,2].

Many of the complex biological effects of GCs depend on activation of the glucocorticoid receptor (GR), a hormone inducible transcription factor [3]. The GR is a class I nuclear receptor (NR) and a modular protein containing three distinct functional regions, a N-terminal activation domain (AD) containing a smaller defined activation function region (AF-1), a central DNA-binding domain, and a C-terminal ligand binding domain that can also contains a ligand-dependant AF region (AF-2) (Figure 2.1). The mechanism of

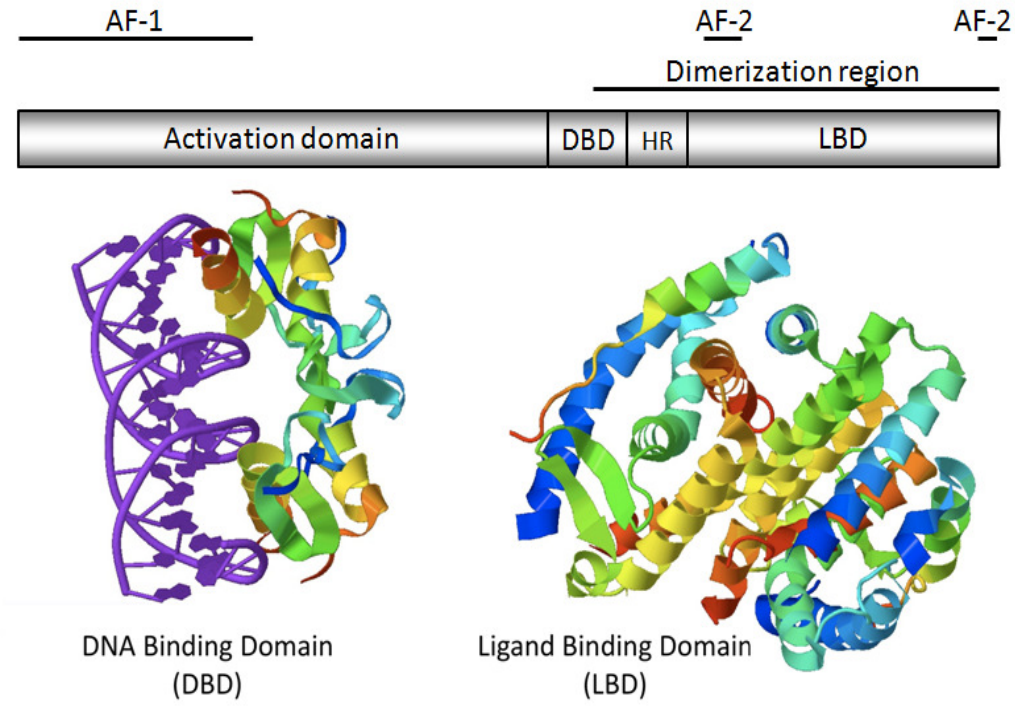


Figure 2.1 The Glucocorticoid receptor domain structure.
The N-terminal activation domain (AD) acts as a ligand independent activation domain interacting with other protein cofactors. DBD is the DNA binding domain separated from the ligand binding domain (LBD) by a hinge region (HR). AF-1 and AF-2 regions are critical for transcriptional activation of GR-regulated target genes. *(Modified image from the Nuclear Receptor Resource http://nrresource.org)*

transcriptional regulation by the GR has long been studied. Upon ligand exposure, the GC-bound GR complex translocates to the nucleus where it recognizes and binds directly to specific DNA sites called GR response elements (GREs). The GR can also bind DNA indirectly through a tethering mechanism involving protein-protein interactions with other transcription factors such as activating protein-1 (AP-1) and nuclear factor kappa B (NF-κB) (reviewed in [4]). After binding DNA, GR can either activate or repress transcription depending on the integration of multiple factors [5]. The net effect of GCs on an individual gene target depends upon other transcription factors present on both the target promoter and/or distal regulatory sites such as enhancers [6,7]. Thus, understanding the full mechanism of GC action requires identifying not only the set of genes bound and regulated by the GR, but also the other transcription regulatory factors that may interact with the GR, and the disperse genomic loci where these interactions occur.

Covalent chromatin modifications to both histones and DNA have significant impacts on chromatin structure, function, and gene regulation, as well as emerging as important mechanisms in NR target gene regulation [8,9]. NRs can affect chromatin by recruiting coactivator or corepressor proteins with chromatin or DNA modifying activities. For example, GR and many other NRs, recruit histone acetyltransferases (HATs) such as CBP, p300, p/CAF and SRC/p160, as well as ATP-dependent chromatin remodelers such as Swi/Snf proteins to initiate gene activation [10-13]. The actions of these protein complexes, by addition of acetyl groups to histones and chromatin decondensation, allow basal transcription machinery and RNA PolII access to target genes [14-17]. Conversely, corepressors such as histone deacetylases (HDAC) and NURD complexes recruited by NRs,

repress transcription by removal of acetylation and chromatin condensation [18]. Together, the dynamics of histone acetylation and deacetylation offer a model for an epigenomic switch between on/off states of hormone target genes. However, the epigenome is extremely complex. Its regulation and associated gene activity is cell-type and gene specific, as well as requiring combinatorial histone codes and multi-layer mechanisms. Therefore, the current model of hormone-induced gene regulation is likely quite incomplete. For example, the most abundant and diverse histone modification in the epigenome is histone methylation. It has an incredibly complex impact on gene regulation, having been functionally linked to activation and repression depending on the genomic location, the methylated residue, or the degree of methylation (one, two, or three methyl groups). Similar to acetylation, histone methylation is regulated by a balance of apposing enzymes known as histone methyltransferases (HMTs) and histone demethylases (HDMs), which add or remove methyl groups, respectively.

The recent discovery of HDMs demonstrated the reversibility of histone methylation but also initiated the now rapid-pace characterization of histone demethylation in epigenetic transcriptional control. LSD1/KDM1A, the first identified histone demethylase, plays an essential role in a broad spectrum of biological processes, including gene regulation, stem cell differentiation, embryonic development, and tumorigenesis [19-22]. Not surprisingly, LSD1's mechanism of action in gene regulation (repression versus activation) is complex. LSD1 was originally characterized as a corepressor with *in vitro* validated demethylase activity against H3K4me1/2, an activity that can be regulated by associated protein such as CoREST and MTA [23,24]. Soon after the discovery of its H3K4me1/2 demethylase activity,

LSD1 was also described as a H3K9 demethylase in the context of the hormone receptors ER and AR, offering one of the first examples of substrate specificity switching (K4me to K9me) by a HDM [25-30]. However, this mechanism remains somewhat controversial given the lack of clear biochemical or structural support for how LSD1 changes substrate specificity to directly demethylate H3K9 [31-33].

Transcriptional control by NRs through distal regulatory sites, such as enhancers, has been implicated as a key mechanism underlying hormone-mediated gene regulation [34-37]. Recent work has begun to define distinctive histone modifications at enhancers that contribute to epigenetic control of enhancer activity. Heintzman et al. first defined an active enhancer chromatin signature as enriched in H3K4me1 and H3K27 acetylation (ac). While studies suggest H3K27ac is a key mark in the transition from inactive or poised enhancer states to fully active [38], growing evidence also supports a critical role for various H3K4 methyl states. H3K4me1 is the most generally accepted H3K4 feature of enhancers. However, H3K4me2 has been implicated as required for DNA binding of the ER and AR pioneer factor FoxA1 [39]. Lastly, a low level of H3K4me3 is also a hallmark of active enhancers distinguishing them from active transcription start sites (TSSs) [38,40]. Thus, an enhancer-specific H3K4 methylation profile for all three methyl-states has been described.

Surprisingly, little is known about cellular, molecular, and biochemical mechanisms regulating changes in H3K4 methylation at enhancers. One study brought to light the importance of LSD1 in enhancer regulation in ESC differentiation where erasure of H3K4me1 by LSD1 at enhancers suppresses pluripotency transcriptional networks [19].

LSD1's dual role in gene repression and activation as well as concerns about bona fide substrate switching by LSD1 leaves open the question of whether LSD1 is a direct H3K9 demethylase, and if not, what other mechanism(s) could account for its coactivator function in hormone-induced gene activation. The emerging roles of H3K4 methylation and LSD1 at gene enhancers as well as the importance of these sites in NR-mediated gene regulation raise the intriguing possibility that LSD1 could play a role in fine-tuning the methyl-states of its well-established substrate, H3K4, during the process of enhancer activation.

Potentially related to this dearth of knowledge surrounding regulation of H3K4 methylation at enhancers is the poorly understood cell-type specificity seen with NR function. GR in particular is the well known for a diversity of physiological effects. The ability of GCs to regulate such wide-ranging processes is in part due to tissue and cell-type specific affects. Transcriptomic analysis of GR action has revealed minimal overlap in GC-responsive expression profiles of different cell types [34,41,42]. One specific example that highlights this specificity well is the affect of GCs on cell survival. In lymphoid cell lineages such as T-cells and monocytes GCs induce apoptosis, contributing to immunosuppression and anti-inflammatory effects [43-45]. However, in other cell types such as liver, osteoblasts, glioma, and lung carcinoma cells GCs have been found to inhibit apoptosis [46-51].

Genome-wide binding studies have also revealed that GR selectively targets a subset of GREs present in the genome [34,37,52-54]. While there are a multitude of sites in the genome with sufficient sequence similarity to the GRE consensus motif, only a small fraction have

been found to be functional in terms of GR recruitment. In addition, DNase I hypersensitivity mapping, which is indicative of open versus closed chromatin states, has shown that the majority (~88%) of GR bound sites are kept in a constitutively accessible state [55]. As the functions of epigenetic mechanisms are elucidated, a clear relationship between histone modification and chromatin state is being established. Together this strongly implicates histone modifiers in regulating a cell-type specific chromatin signature that is laid down prior to and independent of hormone exposure and that likely serves two functions. One is to establish a more permissive chromatin state (ie get target sites "primed and ready") for GR induction; and two, to assist in targeting the GR complex to specific sites in the genome. However, the epigenetic modification and the proteins that regulate those modifications that underlie selective genomic targeting by the GR remains unknown and is a key question in understanding the versatile action of GCs in human physiology.

Here, using a GC-inducible model, we illustrate two new molecular mechanisms where LSD1 contributes to both GR-mediated target gene activation as well as suppression of non-functional GR targets by differentially controlling the H3K4me1/2 states at gene enhancers. By microarray analysis we find that LSD1 is important for activation of most GC-responsive genes but also, that loss of LSD1 unmasks a large number of newly GC-responsive genes. As part of the GR complex LSD1 is recruited to many GR bound enhancer sites. Mechanistically, LSD1 H3K4 demethylase activity is crucial for changes in H3K4me1/2 and enhancer activation in response to GCs, challenging the dogma of LSD1 and NR-mediated gene activation via the elusive H3K9 demethylase activity. Contributing to this mechanism we find that GR modulates LSD1 HDM activity *in vitro* and *in vivo* by inhibiting removal of

H3K4me1 allowing this important enhancer mark to accumulate during activation. Additionally, we find that H3K4me2 is an important mark distinguishing functional GR binding sites prior to hormone exposure. Together these data support two roles for LSD1 in GR-mediated transcriptional regulation. One, in response to GCs GR-LSD1 interaction modulates LSD1 enzymatic activity to regulate the balance between H3K4me1 and me2 methyl-states at enhancers. And two, independent of GR, LSD1 and H3K4me2 contribute selective activation of a specific set of genes.

Results

LSD1 and GR form a stable complex dependent on the GR activation domain.

To begin understanding the role of HDMs in GR-mediated gene regulation we tandem affinity purified (TAP) the GR-associated complex. A stable cell line expressing FLAG- and HA-tagged GR was generated in HeLa suspension cells. Cells were treated with the synthetic GR agonist Dexamethasone (DEX) prior to purification by FLAG and then HA immunoprecipitation. Silver staining of the FLAG and subsequent doubly purified HA sample along with a mock purification are shown in Figure 2.2 A, demonstrating high and specific enrichment. MS/MS analysis of the GR complex identified many proteins and complexes involved in chromatin modification/organization, gene transcription and regulation (Figure 2.2 D). Several key chromatin modifiers were enriched, including components of the Swi/Snf complex, the MLL2 H3K4 HMT complex, several HDACs, JMJD1C (a H3K9/36 HDM), and LSD1 (Figure 2.2 C). To confirm the stable interaction between LSD1 and GR, a third IP using LSD1 antibody was carried out from the HA TAP sample. The

reciprocal purification of GR subcomplexes containing LSD1 demonstrates the stable formation of the GR/LSD1 complex (Figure 2.2 B lane 2, and 2.2 C lanes 6 and 7).

GST pull-down assays using recombinant purified proteins were used to further characterize the LSD1-GR interaction. Several GR truncations were generated (Figure 2.3 A) to determine which region of the GR is most important for binding LSD1. Full length GST-GR pulled down His-LSD1 (Figure 2.3 B, lane 3) indicating the two proteins interact directly. The Gm1-419 truncation, but not Gm420-777, was also able to pulldown LSD1, suggesting the N-terminal activation domain (AD) is necessary and sufficient to bind LSD1. Additional AD truncations showed little binding activity suggesting no single smaller region within the AD is sufficient for strong binding to LSD1 (Figure 2.3 C).

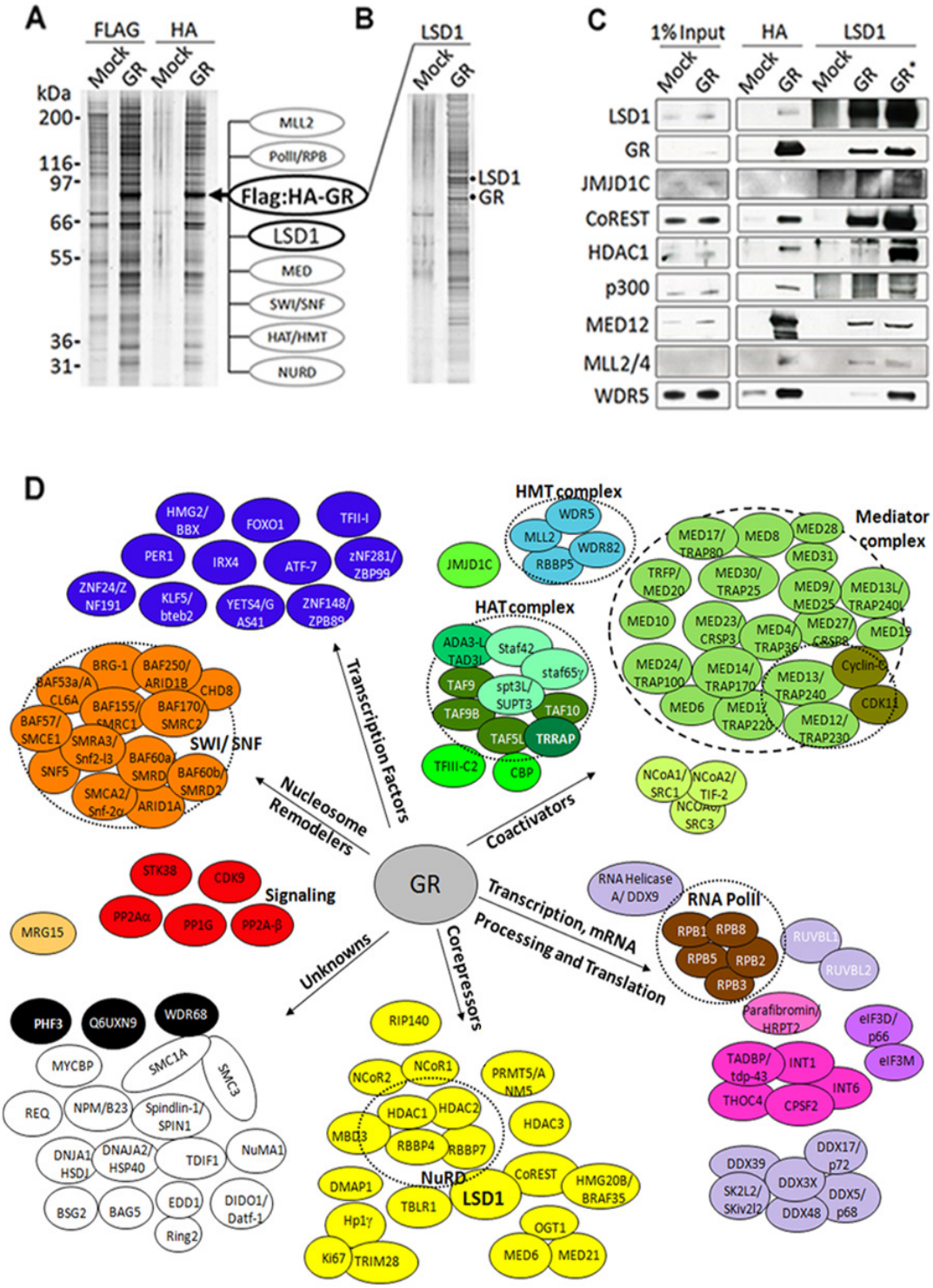


Figure 2.2 LSD1 and GR form a stable complex.

Figure 2.2 (Continued)

GR complex was tandem affinity purified (TAP) from stable cell lines expressing either HA-FLAG vector (Mock) or HA-FLAG-GR (GR). Silver staining of purified complexes shows numerous GR binding partners (A). A portion of the final HA purification was used for a third round of purification against LSD1 to isolate GR-LSD1 complexes (B). IP bait is indicated at the top of the gel, and the vector expressed in the cells above each lane.

Another portion of the HA purification was sent for MS/MS identification of co-purified proteins. GR associated proteins identified by MS/MS were analyzed and grouped based on published literature (D).

Western blots confirmed many proteins identified by MS/MS in the GR and GR-LSD1 complexes (C). IP bait is indicated at the top of the blot, the vector expressed in the cells above each lane, and the antibody probe on the left. "GR*" was TAP against FLAG and then LSD1 to obtain GR-LSD1 complexes through only two purifications instead of three.

LSD1 is recruited to endogenous GR binding sites and DEX-induced genes.

Given that LDS1 is part of the GR complex and directly interacts with GR, we next determined if LSD1 is recruited to endogenous GR binding sites *in vivo*. We preformed chromatin immunoprecipitation coupled with deep sequencing (ChIP-Seq) for GR and LSD1 after DEX or an ethanol vehicle (EtOH) treatment in the lung adenocarcinoma cell line, A549, which is an established model line for studying GR and GC action [34,41]. In total,

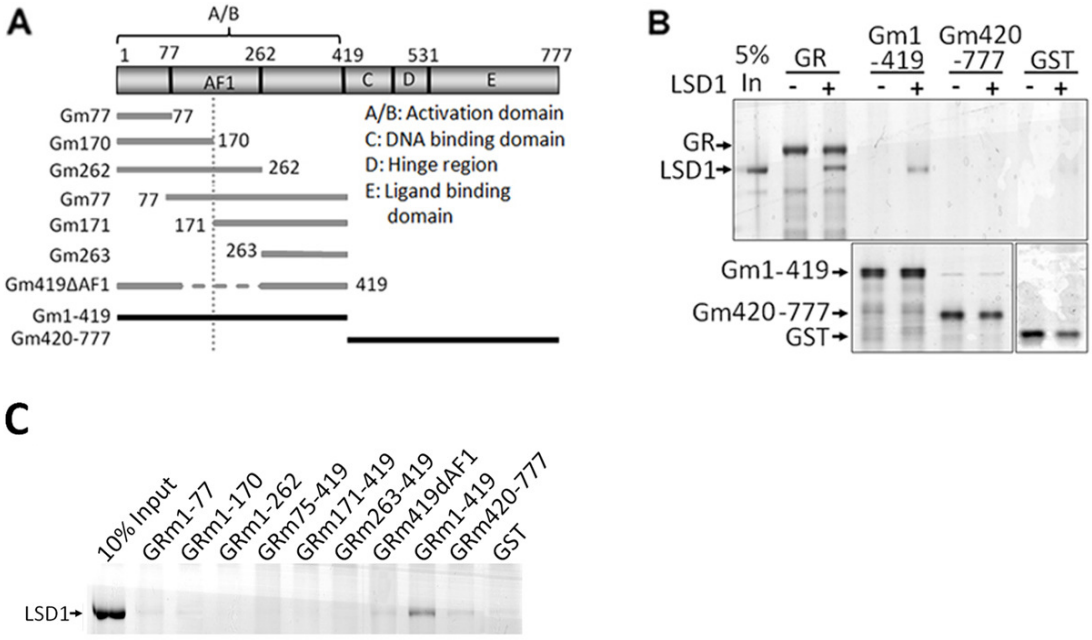


Figure 2.3 LSD1 and GR directly interact via the GR activation domain (AD) (1-419aa) *in vitro*.

Various GR truncations were generated (A) used for GST pulldown and demethylase assays. Silver stain of GST pulldown assays (B-C) showed HIS-LSD1 interacted with full length GR as well as the truncation (Gm1-419) containing the GR AD, but not Gm420-777 which lacks the AD, or further truncations of the AD.

1417 GR peaks were identified after DEX, and as expected, very little GR binding detected after EtOH. GR peaks were analyzed by the motif-calling program MEME, which returned the GRE consensus sequence suggesting the majority of binding sites contain GREs. Only a small percent (3.6%) of GR peaks were found within 2Kb of a known TSS (Figure 2.4 A), consistent with previous NR profiling [34,35].

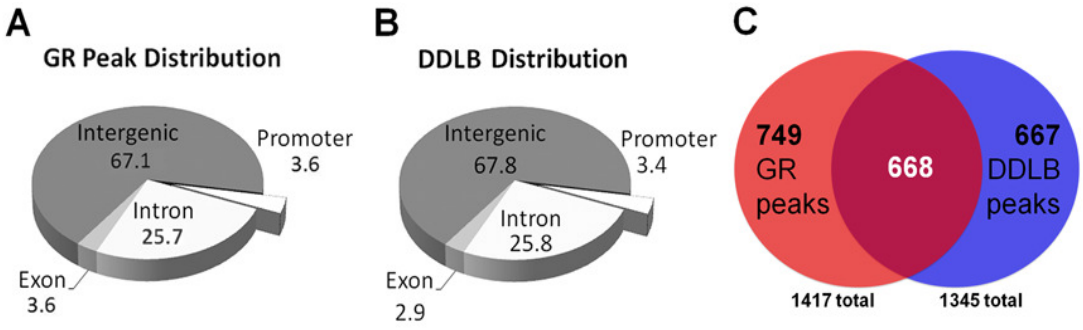


Figure 2.4 DEX-dependent LSD1 recruitment overlaps many endogenous GR binding sites.

ChIP-Seq against LSD1 and GR from A549 cells treated with ethanol (EtOH) or 100nM DEX (DEX) for 2 hours. Genomic distribution of DEX GR peaks (A) and the genomic distribution of DEX-dependant LSD1 binding (DDLB) (B) was very similar and primarily outside of promoters. DDLB sites were defined as by LSD1 DEX/EtOH density ratio ≥ 1.5. Many GR peaks showed DEX-dependent increases in LSD1 binding demonstrated by the overlap between GR and DDLB peaks (C).

LSD1 ChIP-seq on the other hand showed binding throughout the genome in both the EtOH and DEX treated samples, suggesting a broad regulatory role in agreement with previous reports as well [30,56,57]. Therefore, analysis of LSD1 binding focused on differential loci displaying DEX-dependent LSD1 binding (DDLB). These loci were defined as regions with ≥ 1.5-fold higher LSD1 signal density in the DEX sample compared to EtOH. The genome wide distribution of DDLB sites was very similar to GR with only 3.4% occurring at promoters (Figure 2.4 B). To address the question of LSD1 recruitment to GR binding sites, loci showing GR recruitment were compared to DDLB sites finding that 668 GR peaks (47%) showed ≥ 1.5-fold enrichment in LSD1 (Figure 2.4 C). Examples of GR and LSD1 profiles showing increased LSD1 binding at both GR bound regions as well as the TSSs of four DEX activated genes are shown in Figure 2.5.

To validate the genome-wide analysis, we employed ChIP-qPCR to examine GR (Figure 2.6 A) and LSD1 (Figure 2.6 B) at four DEX-activated genes with distal GR bound sites. DEX-induced genes lacking GR at the promoter were selected with the goal of better understanding potential enhancer functions. Specifically, *per1* has well-characterized GRE and GR peak 2Kb upstream of the TSS (labeled 2kGRE). *birc3* has a GR peak 12kb upstream of TSS (labeled 12kGB) and very close to a site identified previously [34]. While to our knowledge GR binding sites or enhancers for *cdkn1c* or *dusp1* have not been characterized our analysis identified the closest peaks at 105Kb downstream of the *cdkn1c* TSS (labeled 105kGB) within the body of a neighboring gene (KCNQ1, not regulated by DEX), and 1Kb upstream of the *dusp1* TSS (labeled 1kGB). An intergenic region (interg) not bound by GR was used as a negative control. The results show significant DEX-dependent recruitment of LSD1 at GR peaks, validating the ChIP-Seq analysis. Moreover, for three of the four genes we also detect significant LSD1 enrichment at the TSS by qPCR. These results are consistent with the presence of LSD1 in the GR complex and strongly support a direct role for LSD1 in regulating GR targets gene activation.

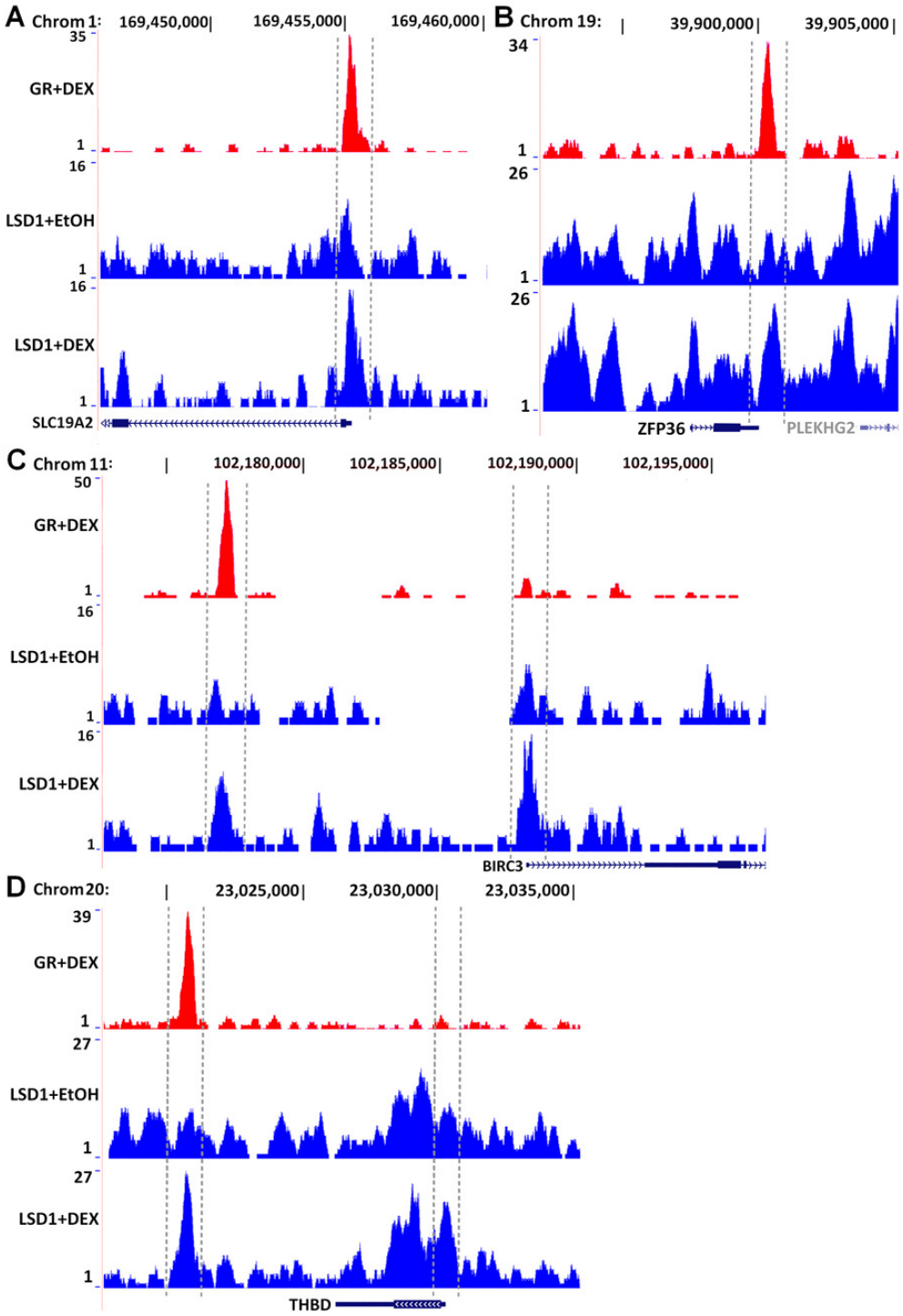


Figure 2.5 Many GR binding sites as well as the promoters of DEX activated genes show DEX-dependent LSD1 recruitment.

Figure 2.5 (Continued)

GR (red) and LSD1 (blue) ChIP-seq profiles from four DEX induced genes *slc19a2* (A), *zfp36* (B), *birc3* (C), *thbd* (D), genes showed increased LSD1 binding after DEX at GR bound sites. In two examples GR binding is close to the TSS (*slc19a2* and *zfp36*) while the other two genes show GR binding over 10Kb from the TSS (*thbd* and *birc3*). LSD1 recruitment is observed at both the distal GR binding sites and the TSS of *birc3 (C)* and *thbd* (D).

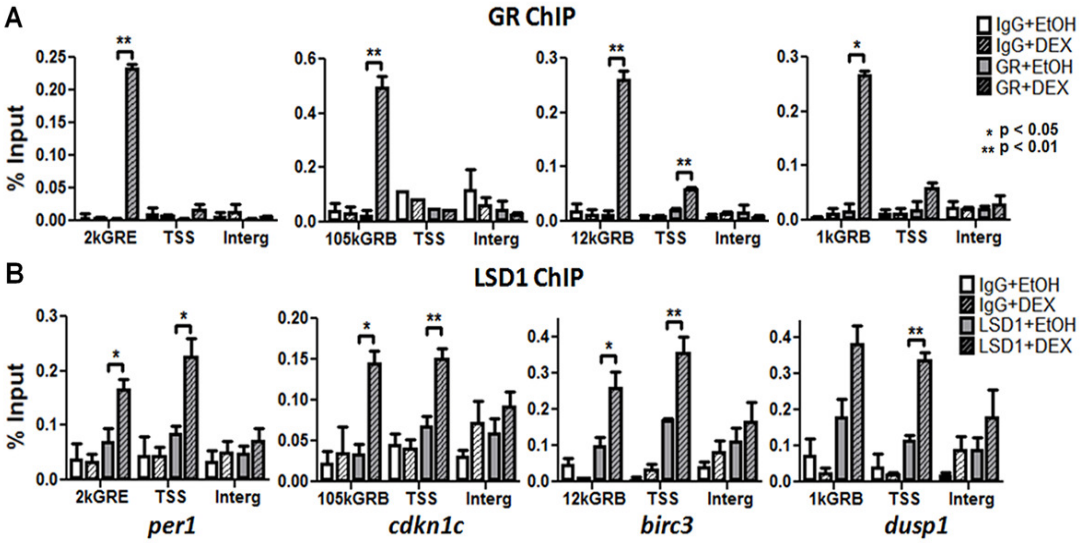


Figure 2.6 ChIP-qPCR confirms DEX-dependent LSD1 recruitment to GR binding sites and promoters of DEX-activated genes.

ChIP against GR (A) or LSD1 (B) from A549 cells showed significant DEX-induced GR and LSD1 recruitment at four sites identified from ChIP-seq. Significant LSD1 recruitment was also observed at the promoters of each DEX-activated gene.

The average of 2-4 biological replicates are plotted ± standard error, p-values from two-tailed t-test.

LSD1 is a key transcriptional coactivator of DEX-responsive genes and functionally linked to GC action.

LSD1 has been shown to act as a coactivator for both ER and AR target genes. To begin addressing whether LSD1 is functionally associated with GR and playing a similar role, we compared global gene expression microarray analysis in A549 cells with or without LSD1 depletion. A lentiviral scramble (SC) control or LSD1-sepcific (L-sh1) shRNA was used to knockdown (KD) LSD1 expression. LSD1 KD was greater than 90% at the protein level and did not affect GR expression (Figure 2.7). Figure 2.8 A diagrams the analysis where DEX-regulated genes were identified by comparing control DEX and EtOH treated samples using a ≥ 2-fold cutoff. LSD1-dependent genes were identified by comparing the level of DEX regulation in L-sh1 to SC, using the formula (DEX/EtOH SC) / (DEX/EtOH L-sh1). Genes with >2-fold difference in activation or repression (i.e. 50% defect in the magnitude of activation or repression in L-sh1) were considered LSD1-dependent.

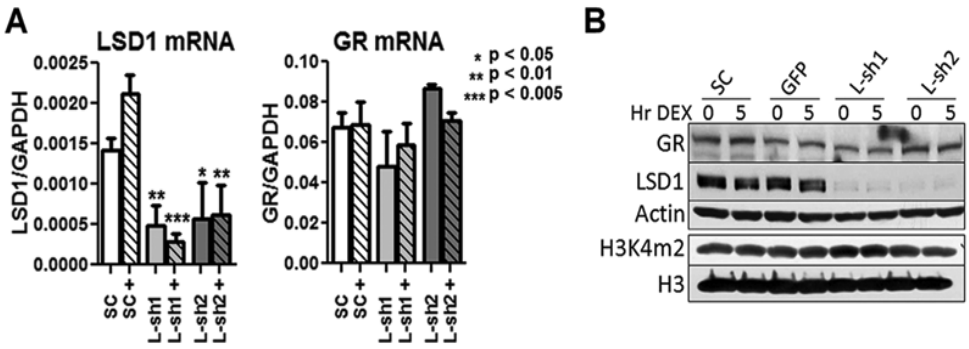


Figure 2.7 LSD1 knockdown by lentiviral shRNA is efficient and does not significantly affect GR expression.

Lentiviral knockdown (KD) in A549 with either scramble control (SC), GFP control (GFP), or two different shRNA constructs against LSD1 (L-sh1, L-sh2) showed significant reductions in LSD1 mRNA (A) and protein (B), but no significant change in GR. The average of three biological replicates is plotted ± standard error, p-values from two-tailed t-test.

Figure 2.8 B shows heatmaps comparing mRNA levels after DEX or EtOH treatment in the SC control and L-sh1 samples. 225 genes were activated and 242 genes were repressed at least 2-fold by DEX in SC. 95 of the 225 activated genes (42%) showed at least a 50% defect in activation in the L-sh1 sample (Group B in Figure 2.8 B), and 123 of the 242 repressed genes (51%) showed at least a 50% defect in repression (Group D in Figure 2.8 B), suggesting that overall about 47% of DEX-regulated genes are LSD1-dependent. Genes in the heatmap are ranked from top to bottom by the ratio of DEX regulation in SC compared to L-sh1 for each group. Thus, the genes at the bottom of Group A or C show a mild defect in activation or repression but did not meet the 50% cut off for classification into the LSD1-dependent Group B or D.

Interestingly, 1,351 genes were also identified as DEX-regulated in the L-sh1 sample but not in the SC, 830 of which were up-regulated (Group E in Figure 2.8 B) and 521 were down-regulated (Group F in Figure 2.8 B). While some these genes appear to be only mildly DEX-regulated (i.e. activated or repressed less than 2-fold), and therefore were not included into Groups A-D, many of these genes do appear to be insensitive to DEX under control conditions. However, they become new DEX targets in the absence of LSD1; suggesting LSD1 is playing a role in repressing or masking these genes from DEX-regulation.

We hypothesized that these genes masked LSD1 are non-specific targets regulated by DEX in other cell types. To test this hypothesis we compared the genes in Groups E and F to DEX-induced expression profiles for four other cell types (3T3-L1 [58], mouse liver cells [59], podocytes [60], and mouse C2C12 myotubes [61]). We found that overall about 12% of the un-masked genes were DEX-regulated in at least one of the four other cell types, with 9.6% from Group E and 16.0% from Group F represented. GCs regulate gene expression in many cell types, very few of which have been experimentally profiled, and this overlap of newly DEX-responsive genes in the absence of LSD1 with previously characterized DEX-responsive genes might suggest a role for LSD1 in maintaining a cell-type specific GC response.

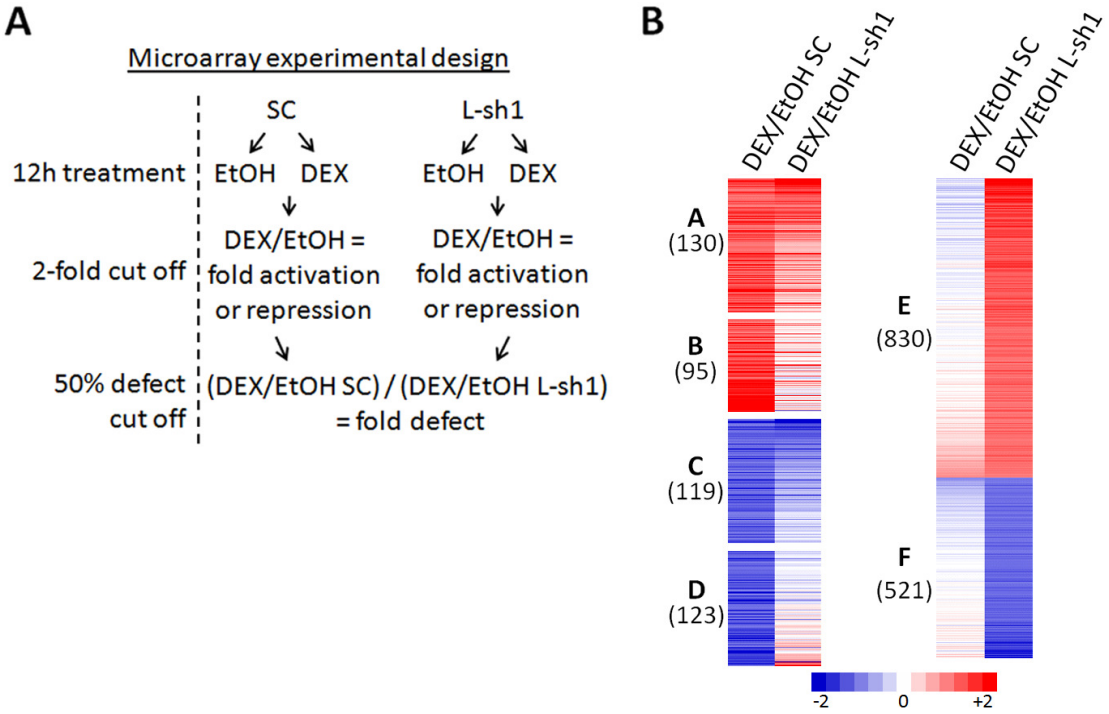


Figure 2.8 LSD1 is a critical transcriptional coactivator of DEX-responsive genes. Microarray analysis following lentiviral KD in A549 with either scramble control shRNA (SC) or LSD1 shRNA (L-sh1) was used to first identify DEX-regulated gene targets by at least a 2-fold change in mRNA, and second, identify LSD1-dependent targets defined by defective up or down regulation by at least 50% compared to SC (A). The heatmap display (B) of DEX and LSD1-regulated genes are grouped based on these criteria. Group A are 2-fold up-regulated in SC and up-regulated in L-sh1 to similar degree. Group B are also up-regulated by DEX 2-fold in SC, but the magnitude of activation after LSD1 KD is reduced by at least 50%, therefore we consider these gene to be DEX-induced and LSD1-dependent. Group C are 2-fold down-regulated in SC and to a similar degree in L-sh1. Group D are also 2-fold down-regulated in SC, but the magnitude of repression after LSD1 KD is reduced by at least 50%, therefore we consider these genes to be DEX-repressed and LSD1-dependent. Groups E and F do not meet the 2-fold cut off for DEX-regulation in SC, but are at least 2-fold up or down regulated in L-sh1. All of the genes in the heatmap are ranked based on the ratio of DEX regulation in SC compared to L-sh1.

Many of the DEX-regulated, LSD1-dependent targets were confirmed with a second shRNA (L-sh2) by RT-qPCR (Figure 2.9 A and C). To further validate LSD1 as a critical coactivator

for GR, we used the MDA-kb2 cell line containing a stably integrated, DEX-inducible MMTV:Luciferase construct. LSD1 KD severely impaired DEX induced luciferase expression (Figure 2.9B).

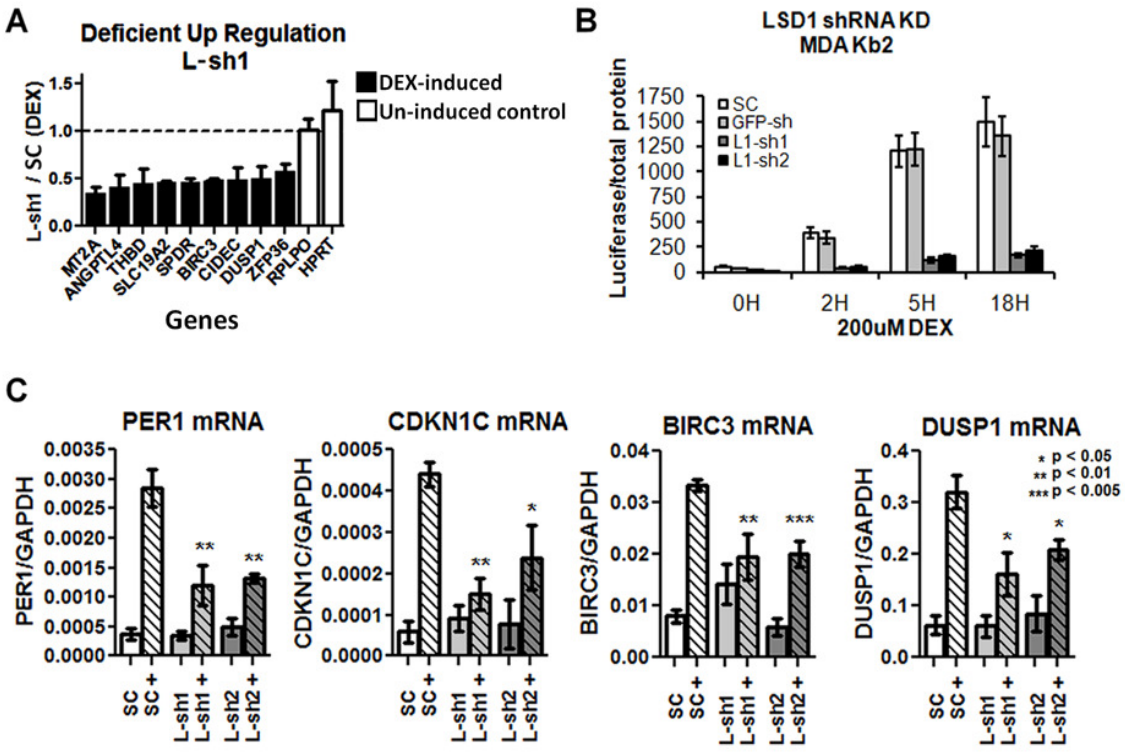


Figure 2.9 LSD1 knock down significantly impairs activation of many GC-target genes. RT-qPCR confirmed many DEX-induced targets require LSD1 for full activation (A) in A549 cells and using a second LSD1 specific shRNA (L-sh2) (C). Similarly, a well-known GC-induced promoter, the MMTV driving a luciferase reporter gene, also required LSD1 for full DEX-activation (B). The average of 3-4 biological replicates is plotted ± standard error, p-values from two-tailed t-test.

LSD1 coactivator function is important for DEX-mediated affects on proliferation, apoptosis, and adipogenesis.

If LSD1 is an important transcriptional co-regulator for GR, this role should translate downstream of transcription to impact the biological effects of GCs. To begin to understand which biological processes might be affected by LSD1 we used the DAVID functional annotation tool to carry out gene ontology analysis of DEX- induced and LSD1-regulated gene groups. The top p-value scores in both groups were negative regulators of apoptosis, consistent with previous reports of apoptosis suppression by DEX in A549 but also importantly implicating LSD1 as a key regulator of that suppression (Figure 2.10). Proliferation, another known DEX-regulated processes, was also significantly enriched as a top DEX-LSD1-regulated group (Figure 2.10 B, black bars). Interestingly, inflammation, a well-known DEX-regulated processes did not show up in the DEX-LSD1-regulated list (Figure 2.10 A, black bars). Together this analysis suggests LSD1 is important for specific sets of GC-regulated biological processes such as proliferation and survival, but may not play a critical role in others, such as inflammation.

To validate these top biological hits through functional studies we examined the role of LSD1 in GR-mediated affects on apoptosis and proliferation in A549. Indeed LSD1 is an important coregulator of DEX-mediated apoptosis and proliferation suppression as demonstrated by TUNEL and CCK8 cell viability assays (Figure 2.11).

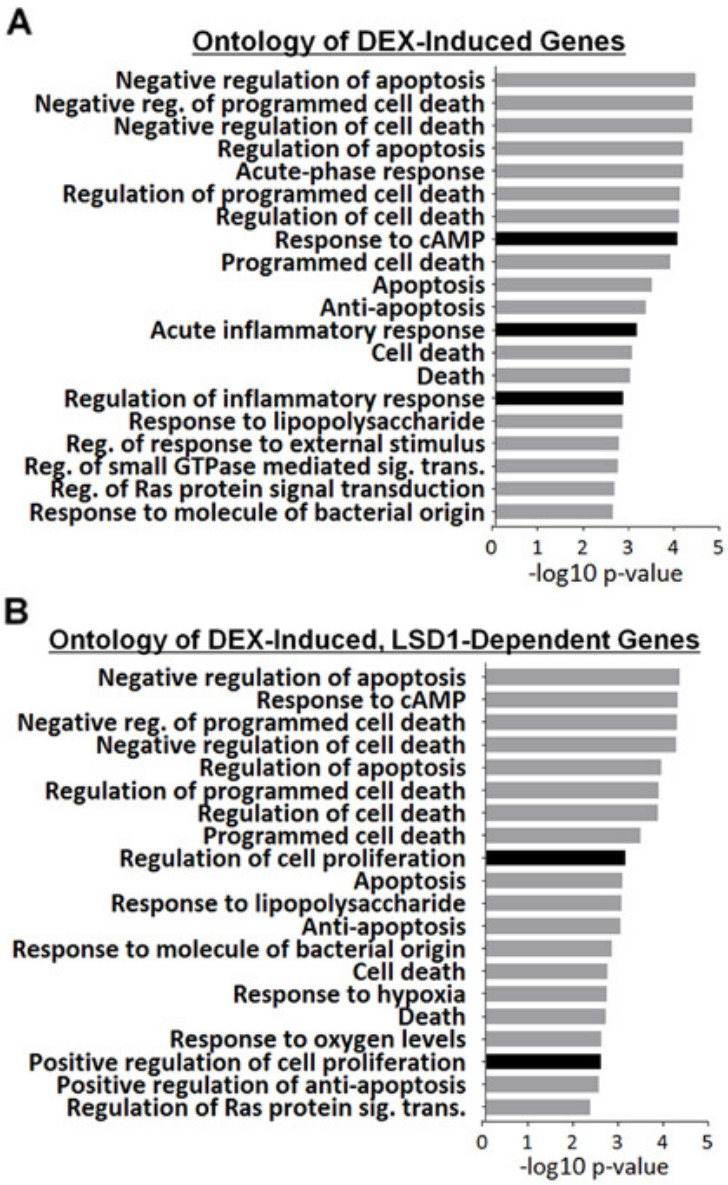


Figure 2.10 LSD1 coactivator function for DEX-responsive genes is linked to GC biological action.

GO analysis of DEX-induced genes from microarray. (A) Top 20 hits of DEX-induced genes ranked by p-value.

(B) Top 20 hits of DEX-induced and LSD1-dependent ranked by p-value. Black bars highlight categories of interest that are unique to either (A) or (B).

GCs also play important roles in development and cellular differentiation including adipogenesis. Stromal cells isolated from bone marrow (bMSC) and grown in tissue culture can be treated with a cocktail of signaling factors including DEX to promote differentiation into mature adipocytes. While an accurate chronology of all the steps involved in adipogenesis have yet to be elucidated growth arrest is known to be an early event followed by the expression of several key regulators including C/EBPβ, PPARγ, followed by C/EBPα leading to cellular remodeling and expression of many important adipocyte specific genes such as Leptin (reviewed in [62]). The cell line A549 is derived from lung tissue and therefore lacks the ability to differentiate into adipocytes. However, two gene targets known to be involved in adipgenesis (CIDEC and C/EBPβ) were found to be regulated by GCs and LSD1 in our microarray, suggesting a possible role for LSD1 is GC-mediated adipogenesis.

To test this possibility mouse bMSCs were isolated from mice containing a LSD1 CRE cassette. Infection with a CRE containing adenovirus deleted the LSD1 gene. bMSCs were infected with either a CRE or EGFP control and grown in differentiation media containing for 14 days before fixation and Oil Red staining. RNA was harvested at day 0 (D0) and day 14 (D14) for RT-qPCR analysis. Oil Red staining at D14 showed many fewer cells with visible accumulation of lipid droplets in the CRE infected cells suggesting loss of LSD1 blocks adipogenesis (Figure 2.12 A). RT-qPCR analysis showed disrupted expression of several key adipogenic factors (Figure 2.12 B). Although, C/EBPβ was up regulated similarly in both EGFP and CRE infected cells, PPARγ and several proteins expressed in mature adipose were significantly reduced in the CRE sample.

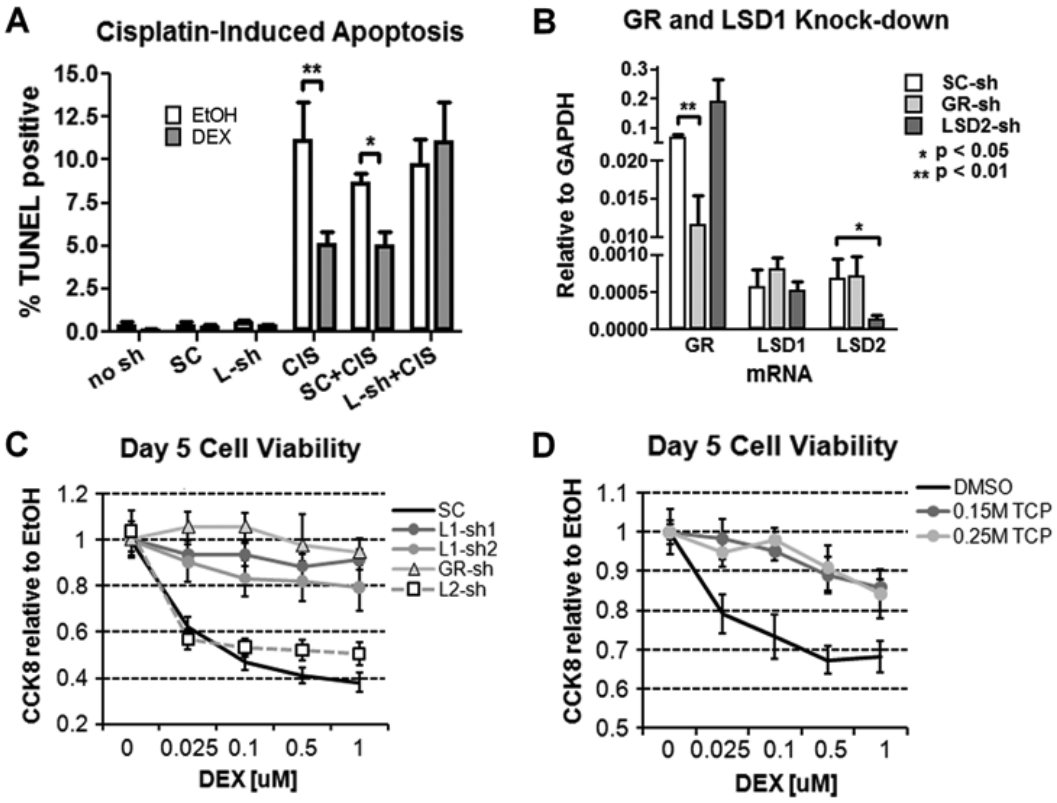


Figure 2.11 LSD1 coactivator function is important for GC anti-apoptotic and anti-proliferative effects in A549.

TUNEL staining was used to measure apoptosis in A549 treated with cisplatin (CIS). Cells treated with 0.5μM DEX with 50μM CIS were significantly resistant to apoptosis compared to co-treatment with EtOH, and KD of LSD1 abolished this resistance (A). CCK8 activity was used to measure proliferation in A549 treated with DEX for 5 days following KD with either a SC control, GR shRNA (GR-sh), LSD1 shRNA (L-sh1 or L-sh2), or LSD1 shRNA (L-sh2). The GR and LSD2 shRNAs were highly specific and did not affect LSD1 expression (B). DEX treatment significantly reduced proliferation (C and D, black line), and KD of GR or LSD1 rendered the cells resistant to proliferation inhibition while KD of LSD2 had no effect. TCP treatment to inhibit LSD1 activity also rendered cells resistant to DEX (D). The ratio of CCK8 activity in DEX vs. EtOH was calculated for each experiment. Average of 3-5 biological replicates is plotted for all, ± standard error, p-values from two-tailed t-test. P-values for GR-sh, or either L-sh compared to SC in (C)

Figure 2.11 (Continued)
and DMSO compared to TCP in (D) were ≥ 0.05 for all DEX concentrations. The average of 3-6 biological replicates are plotted ± standard error, p-values from two-tailed t-test.

LSD1 does not significantly impact DEX inhibition of IL8 expression or release from A549 cells.

GCs are well-known to inhibit inflammation and a simple example of this can be observed using A549 cells. Treatment with TNFα induces A549 cells to secrete the pro-inflammatory factor IL8, and co-treatment with DEX inhibits TNFα-induced IL8 release [63,64]. Because anti-inflammatory affects are one of the most important clinical applications of GCs it was important to test LSD1's role. We first examined IL8 mRNA levels by RT-qPCR and show that DEX treatment represses IL8 expression (Figure 2.13 A) and repression is unaffected by LSD1 KD (Figure 2.13 B). LSD1 and GR ChIP-seq data did not reveal any significant GR or LSD1 recruitment to the IL8 gene (data not shown), suggesting indirect repression in these cells.

DUSP1 is suggested to be an anti-inflammatory regulator downstream of GCs via inhibition of MAPK signaling [65-69]. DUSP1 mRNA increases after DEX treatment (Figure 2.9 C) and was affected by LSD1 KD. Taken together LSD1 does not appear to play a role in regulating IL8 repression in response to DEX, but could potentially affect IL8 release indirectly through regulation of *dusp1*. Unable to predict the affect of LSD1 loss on IL8, secretion was measured directly by treating A549 cells with TNFα to induce secretion detected by ELISA on the media. Figure 2.13 B shows that treatment with DEX reduced the basal levels of IL8

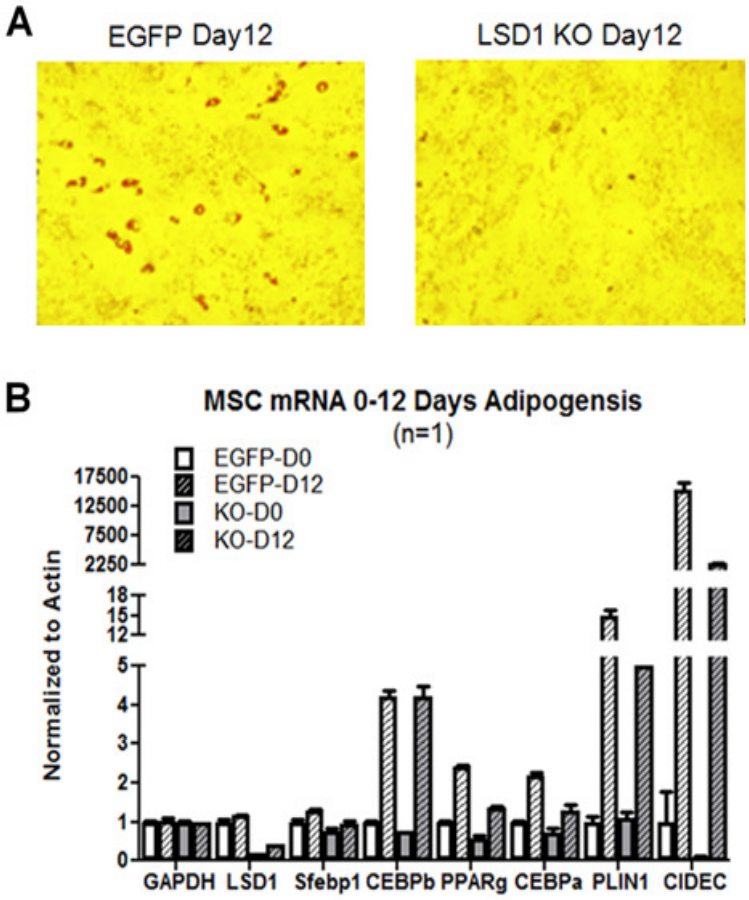


Figure 2.12 LSD1 is important for bone marrow derived stromal cell (bMSC) adipogenesis.

Stromal cells were isolated from CRE-LSD1 mice and infected with either a control (EGFP) or CRE containing virus to delete LSD1 (LSD1 KO). Oil red staining of control and KO bMSCs after 14 days of growth in adipogenic differentiation media showed reduced lipid accumulation in the LSD1 KO cells (A). RT-qPCR expression analysis also showed defective up regulation of several key adipogenic regulators (B). The average of three qPCR replicates from one biological sample is plotted ± standard error.

released as well as the TNFα-induced levels. KD of LSD1 did not affect DEX-inhibition of either the basal or TNFα-induced levels, suggesting LSD1 does not play a significant role in IL8 release in response to TNFα.

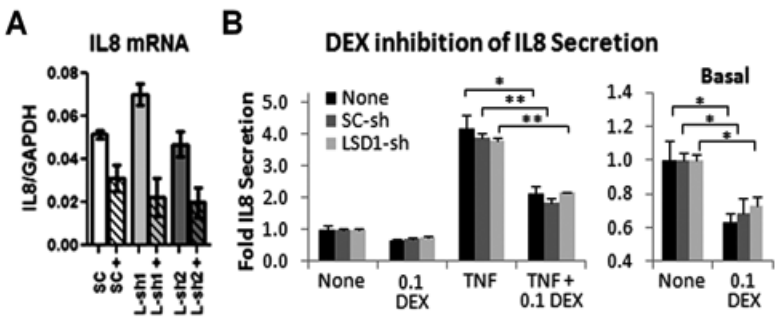


Figure 2.13 LSD1 coactivator function for DEX-responsive genes is linked to GC biological action.
(A) RT-qPCR of IL8 mRNA. (B) IL8 ELISA detecting IL8 secretion by A549 cells into growth media. Cells were treated with 2 ng/mL of TNFα for 24 hours after pretreatment for 6 hours with 100nM DEX (0.1 DEX) and/or viral shRNA for 5 days (dark gray bars are scramble, light gray are L-sh1).
The average of three biological replicates is plotted ± standard error, p-values from two-tailed t-test.

Together these data suggest that LSD1 plays a significant role in DEX-mediated gene regulation that translates from the level of transcription out to the cellular effects of DEX that result from those changes in gene expression.

LSD1 HDM activity is critical for coactivation of DEX-responsive genes.

LSD1 functions primary as is a histone demethylase, making its role in GR target gene regulation likely to involve this activity. Tranylcypromine (TCP) is an amine oxidase inhibitor used to inhibit LSD1 as well as LSD2, but does not inhibit JMJC HDMs. To examine the role of LSD1's demethylase activity, DEX-induced luciferase activity in MDA Kb.2 cells (Figure 2.14), and mRNA levels of endogenous GR targets were measure following TCP treatment (Figure 2.15 A). Both DEX-induced luciferase activity and A549 gene expression

were inhibited by TCP in a dose-dependent manner. To rule out a role for LSD2 as well as off-target effects of TCP, a rescue experiment was performed in A549. Cells were infect with lentivirus carrying a siRNA construct specific for wild-type LSD1 (Li), or vectors containing Li as well as the cDNA sequence for LSD1 with a silent mutation making it resistant to Li. This allows for KD of endogenous LSD1 and simultaneous rescue with the co-expressed LSD1 cDNA. Three different LSD1 cDNA constructs were used, wild-type LSD1 (WT), or two catalytically inactive mutants one with a mutation in the FAD domain (M1) and one with a mutation in the amine oxidase domain (M4). Expression levels of each construct as well as KD efficiency by Li are shown in Figure 2.15 B and C. Similar to L-sh1/2, Li prevented full DEX-induction (Figure 2.15 D), and co-expression of WT rescued this affect bringing expression similar to SC. M1 and M4 did not rescue leaving mRNA levels similar to Li demonstrating the necessity of LSD1 enzymatic activity for coactivator function. Together these experiments strongly support a critical role for LSD1 histone demethylase activity in gene activation by GCs.

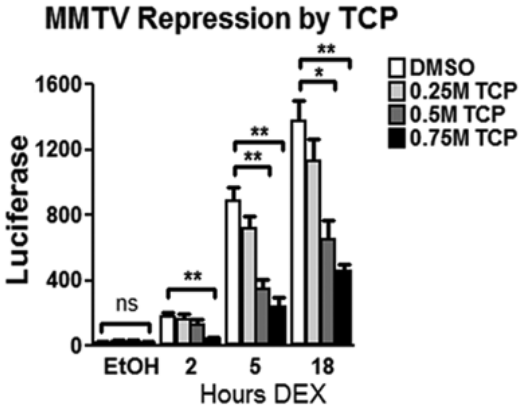


Figure 2.14 LSD1 enzymatic activity is critical for GC-induction of MMTV promoter.
The KDM1 inhibitor TCP antagonized DEX induction of the MMTV:Luciferase construct stably integrated in the MDA Kb.2 cell line. Average of three biological replicates is plotted ± standard error, p-values from two-tailed t-test.

DEX-induction differentially alters H3K4me1/2/3 at GR-target promoters and putative enhancers promoting activation.

To understand the epigenetic mechanism underlying LSD1 coactivator function for GR, we first determine by ChIP-qPCR the status of several histone modifications on GR binding sites and target gene promoters before and after DEX treatment. H3K4me3, an epigenetic mark of gene activation, increased significantly at the TSS of all four genes (*per1*, *cdkn1c*, *birc3*, *dusp1*) as expected for DEX-induction (Figure 2.16 A). While H3K4me2 is also generally considered an activating mark, we surprisingly observed decreased levels at the TSS of all four genes (Figure 2.16 B). In contrast to the dramatic changes in H3K4me2/3, H3K4me1, a recently defined enhancer mark, showed no significant change at the TSS (Figure 2.16 C).

However, this DEX-induced H3K4 profile at the TSS differed from GR bound putative enhancer sites (herein referred to as GB enhancers). Three of the four GB enhancers showed relatively high H3K4me2 levels prior to DEX which, similar to the TSS, decreased upon induction (Figure 2.16 B). Distinct from the TSS, this decrease coincided with a significant increase in H3K4me1 (Figure 2.16 C) while H3K4me3 was depleted and remained so after DEX (Figure 2.16 A). The H3K4me1/3 profile at GBs is consistent with enhancer function, and further supported by observed DEX-induced increases in H3K27ac (Figure 2.16 D) another established mark of active enhancers.

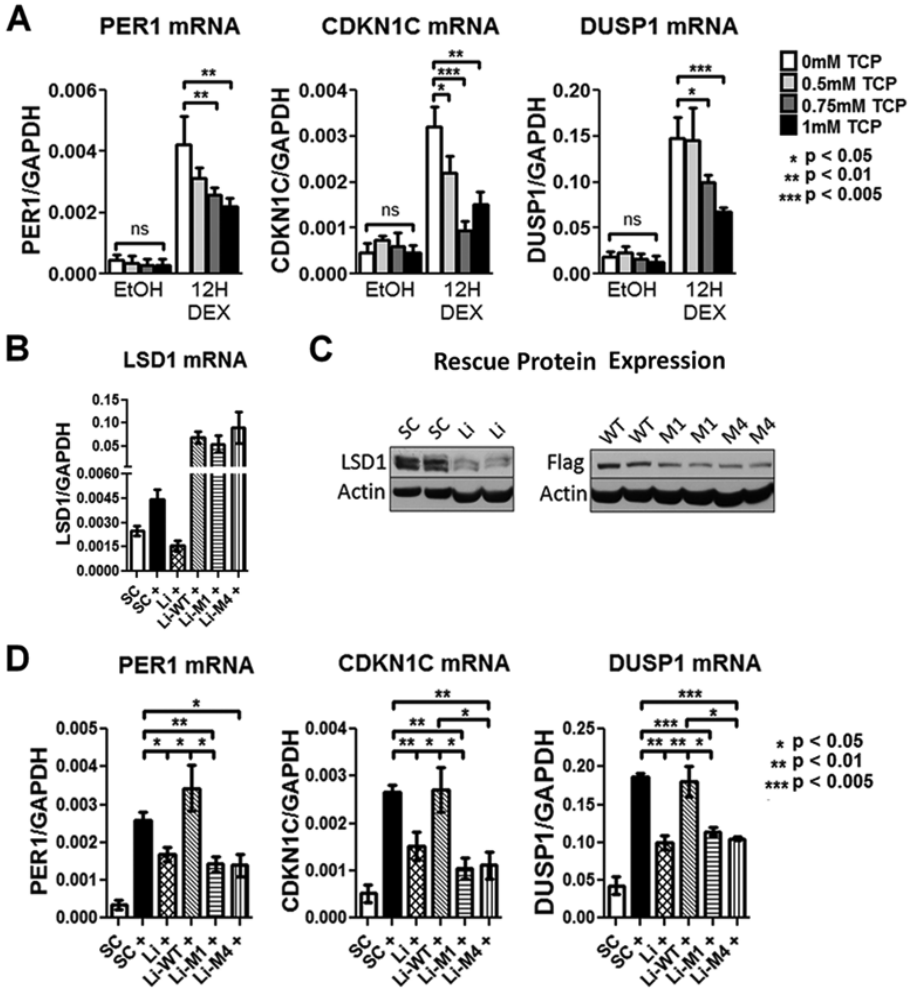


Figure 2.15 LSD1 is a critical transcriptional coactivator of DEX-responsive genes depending on its demethylase activity.

(A) RT-qPCR of several microarray indentified DEX-induced, LSD1-dependent genes after similar TCP treatment in A549. (B) RT-qPCR of LSD1 mRNA after lentiviral infection containing a LSD1 RNAi construct (Li), or Li with cDNA constructs on the same plasmid for wild type LSD1 (WT) or two catalytic mutants (M1 or M4). Average of 3-4 biological replicates is plotted ± standard error, p-values from two-tailed t-test.

Given LSD1's proposed role as a H3K9 demethylase in ER and AR target gene activation we also examined the levels H3K9me2 (Figure 2.16 E), and found most sites both GB enhancers and the TSS did not show significant changes in H3K9me2 after DEX. Two exceptions were the GB nearest *birc3* and the TSS of *dusp1* which both showed a DEX-induced decrease in H3K9me2. Most other sites examined showed low levels of H3K9me2 relative to the intergenic control suggesting that this mark is low prior to hormone treatment and therefore unlikely to be further reduced and may therefore not play a major role in regulating these gene targets. The observation of H3K9me2 loss at the birc3 GB enhancer may indicate that a subset of GB enhancers are more significantly silenced prior to hormone exposure [39]. Taken together the data shows a shift in the H3K4 methylation profile away from me2 and toward me3 at the TSS and toward me1 at GBs enhancers, with only minor contributions from H3K9me2 loss.

To validate observed changes in H3K4me2/1 as a general phenomenon, we examined genome-wide changes by ChIP-Seq. H3K4me1 density measured across all GR peaks and averaged showed a significant increase after DEX (Figure 2.17 A, p=1.21623e-16 calculated from 2-tailed t-test of densities summed across GR peaks), whereas H3K4me2 showed a significant decrease (Figure 2.17 B, p=1.92442e-5 calculated from 2-tailed t-test of densities summed across GR peaks). Similar to the qPCR results, global analysis show that H3K4me2 levels also decreased at the promoters of genes identified by microarray as DEX-induced and LSD1-dependant, while H3K4me1 remained low (Figure 2.17 C). Taken together the data confirms a trend shifting the H3K4 methylation profile away from me2, and toward me3 at the TSS, and toward me1 at GB enhancers.

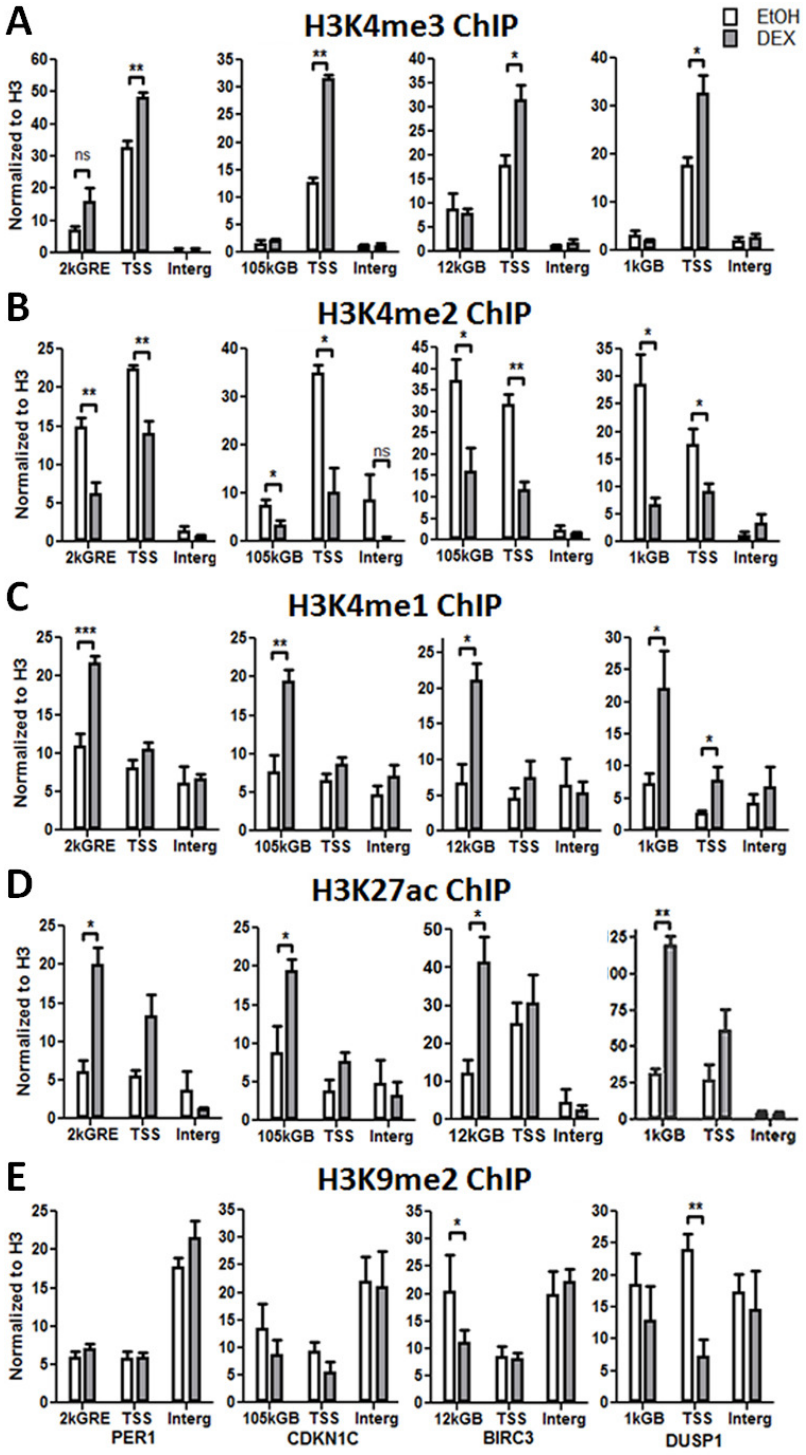


Figure 2.16 DEX-induction changes the H3K4 methylation profile at both enhancers and promoters.

Figure 2.16 (Continued)
100nM DEX treatment for two hours resulted in significant increases in H3K4me3 at the TSS of four DEX-induced genes (A). H3K4me2 levels were significantly reduced at GR bound putative enhancers (GB enhancers) and TSSs (B), and H3K4me1 was significantly increased at GB enhancers (C). H3K27ac was also increased at GB enhancers (D), while H3K9me2 showed mostly low and unchanging levels (E). Average of 2-6 biological replicates is plotted ± standard error, p-values from two-tailed t-test.

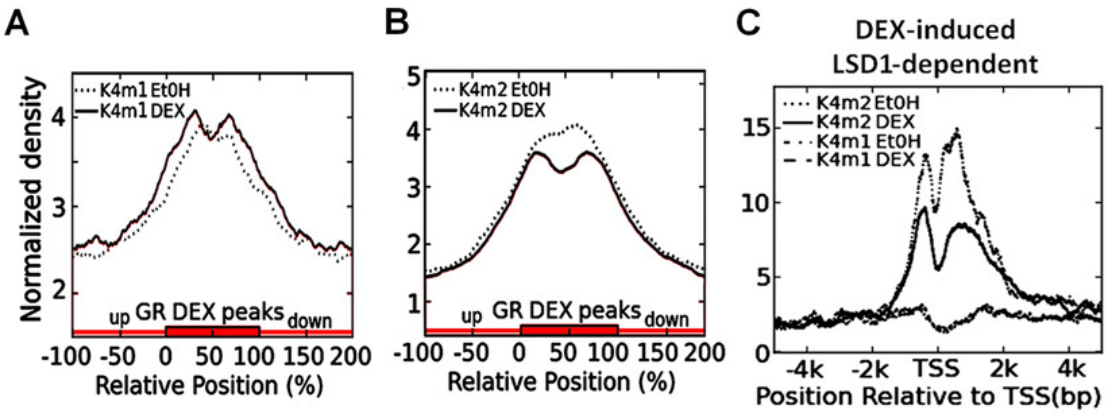


Figure 2.17 Genome-wide analysis confirms trends in H3K4me2/1.
The density of H3K4me1 and me 2 were averaged over all GR ChIP-Seq peaks showing a trend toward increased H3K4me1 (A) and decreased H3K4me2 (B). Loss of H3K4me2 was also observed around the TSS of genes that are DEX-induced and LSD1-dependent as identified microarray.

Strikingly, this analysis also suggests that a high level of H3K4me2 at GR binding sites prior to DEX induction is a prominent epigenetic feature of LSD1-occupied GBs. To better understand the relationship between H3K4me1/2 prior to GR/LSD1 recruitment, a heatmap of LSD1 density on GR peaks after DEX was compared to heatmaps of H3K4me1

and me2 after EtOH treatment, which is representative of the levels prior to DEX (Figure 2.18 A). Loci were ranked by LSD1 density from high to low (top to bottom) revealing a markedly similar pattern in H3K4me2 of high to low density, and a similar but less distinct pattern for H3K4me1. To measure the correlation between LSD1 and H3K4me2, average densities were plotted generating a Pearson correlation coefficient of 0.4587 (Figure 2.18 B). These analyses suggest LSD1 is preferentially targeted to GBs with preexisting high levels H3K4me2. Subsequent H3K4me1 and H3K27ac enrichment (Figure 2.16 C and D) suggests H3K4me2 at GBs may be an important early epigenetic signal distinguishing enhancers ready to be converted to more activate states.

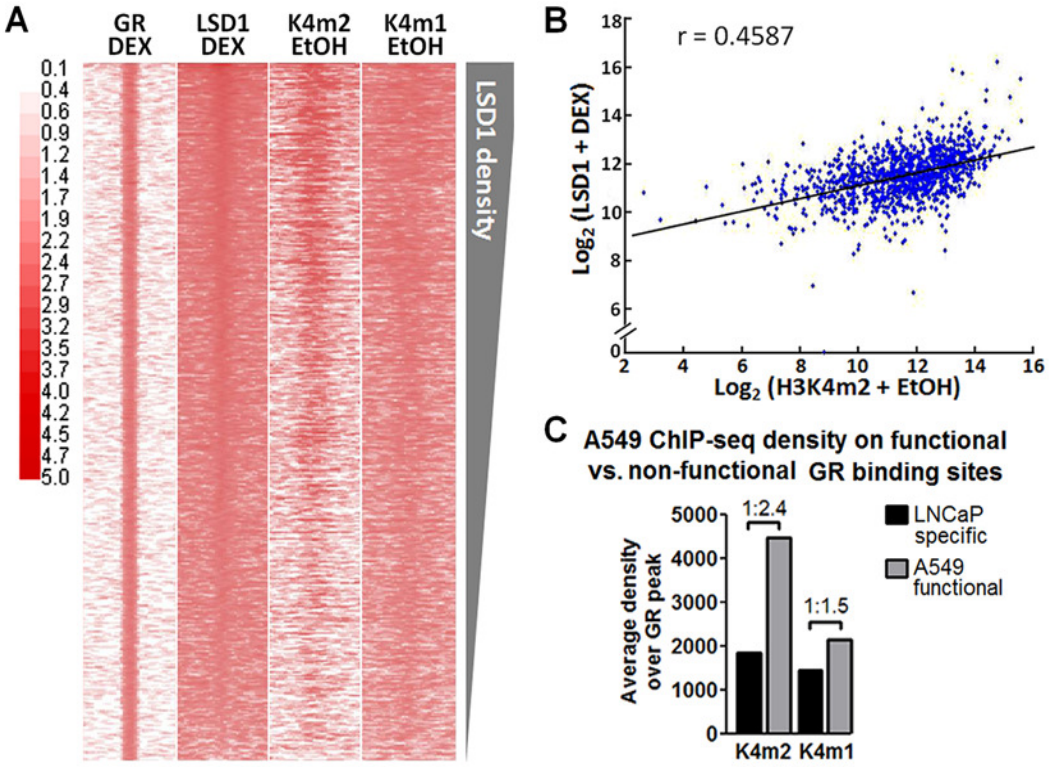


Figure 2.18 H3K4me2 is enriched at functional GR binding sites prior to GR and LSD1 recruitment.

Figure 2.18 (Continued)

Heatmaps of GR and LSD1 binding after DEX, and H3K4me1 and me2 prior to DEX (EtOH) showed a correlation between LSD1 binding after DEX and H3K4me2 levels prior to DEX (A). Plotting the density on GR peaks of LSD1 after DEX and H3K4me2 before DEX showed a Pearson correlation coefficient of 0.4587 (B). GR ChIP-Seq data from A549 was compared to published data in LNCaP cells. GR binding sites unique to LNCaP cells were considered LNCaP specific. The density of H3K4me2 and H3K4me1 was averaged on GR peaks from A549 (A549 function) and compared to LNCaP specific sites showing higher levels of H3K4me2 at functional versus non-functional sites (C).

This led us to question whether H3K4me2 could be a distinguishing mark of functional enhancers prior to activation. Both enhancers and NR binding sites are known to be highly cell-type specific, making predictions of functionality based solely on DNA sequence difficult. To investigate the relationship between H3K4me1/2 and functional regulatory sites, as defined by GR binding, we compared our GR ChIP-Seq in A549 to another cell type. GR binding sites in LNCaP cells were obtained from a publicly available dataset [70] and compared to A549. LNCaP-specific sites were identified by excluding sites bound by GR in both cell lines. These LNCaP-specific sites represent potential GR binding sites given the demonstrated capacity to recruit GR in LNCaP cells, but which we consider non-functional in A549 due to the lack of GR binding in our ChIP-Seq. Once non-functional GR binding sites were identified we compared the levels of H3K4me1/2 at function versus non-functional sites in the A549 EtOH sample (Figure 2.18 C). Quantifying each mark prior to DEX-induction revealed that H3K4me2 was much higher (2.4-fold) at functional GR binding sites compared to non-functional sites prior to DEX, while H3K4me1 was only 1.5-fold higher.

The greater enrichment of H3K4me2 over H3K4me1 prior to stimulation suggests H3K4me2 is a more prominent marker pre-determining GR binding and supporting a role as an early enhancer mark. These studies support a significant role for H3K4 methylation in GR-mediated target gene activation and raise a critical question as to how the H3K4 methyl code is regulated and the potential role of GR, LSD1, and their interactions in that regulation.

GR inhibits LSD1 H3K4, in particular H3K4me1, demethylase activity *in vitro*.

To investigate the enzymatic nature of LSD1 in GR-mediated gene activation, we examine the demethylase activity of the purified GR complex using assay conditions favoring FAD dependent histone demethylation [71]. Surprisingly, unlike previously reported androgen receptor (AR) containing LSD1 complexes showing H3K9me2 demethylase activity [25,26,28], the GR complex containing LSD1 showed no detectible HDM activity against H3K4 or H3K9, suggesting inhibition by a factor(s) in the complex (Figure 2.19 A). Given the direct interaction between GR and LSD1 (Figure 2.3 B), we hypothesized that GR was a likely candidate and used an *in vitro* biochemical approach to test this possibility.

Purified recombinant GR was added to demethylase assays to determine the affect on LSD1 activity. Increasing amounts of GR inhibit LSD1 demethylation of H3K4me1/2 (Figure 2.19 B). This affect was ligand independent (Figure 2.19 B, lane 4) consistent with the finding that LSD1 and GR interact independent of the ligand-binding domain (Figure 2.3 B). DEX in the presence of boiled GR or GST protein had no effect (Figure 2.19 B, lanes 5 and 6). Again unlike the effect of AR or ER on LSD1 activity, no appreciable H3K9me2 demethylation was

detected under any conditions, suggesting that LSD1 was not functioning as a coactivator through direct H3K9 demethylation in the GR complex. Similar inhibitory effects of GR on LSD1 H3K4 demethylase activity were obtained in nucleosome assays which contain CoREST, an LSD1 cofactor required for efficient demethylation of nucleosomes, a more physiologically relevant LSD1 substrate (Figure 2.19 C). GR truncations were again used to test which domain of GR is responsible for inhibiting LSD1 (Figure 2.3 A). Consistent with the GST pulldowns, the AD (1-419aa) was required and sufficient for effective inhibition (Figure 2.19 D).

To assess the affect of GR on LSD1 substrate specificity, densitometry quantification of multiple experiments was used to calculate the fold inhibition on each mark. Inhibition of LSD1 activity against H3K4me2 nucleosomes was modest while inhibition of H3K4me1 loss was significantly stronger (Figure 2.19 E). A similar trend was observed with stronger inhibition of H3K4me1 over me2 demethylation on histone substrates and with GR truncations (Figure 2.19 F). Together these data provide an *in vitro* biochemical explanation for the decrease of H3K4me2 and concurrent accumulation of H3K4me1 at sites co-occupied by LSD1 and GR.

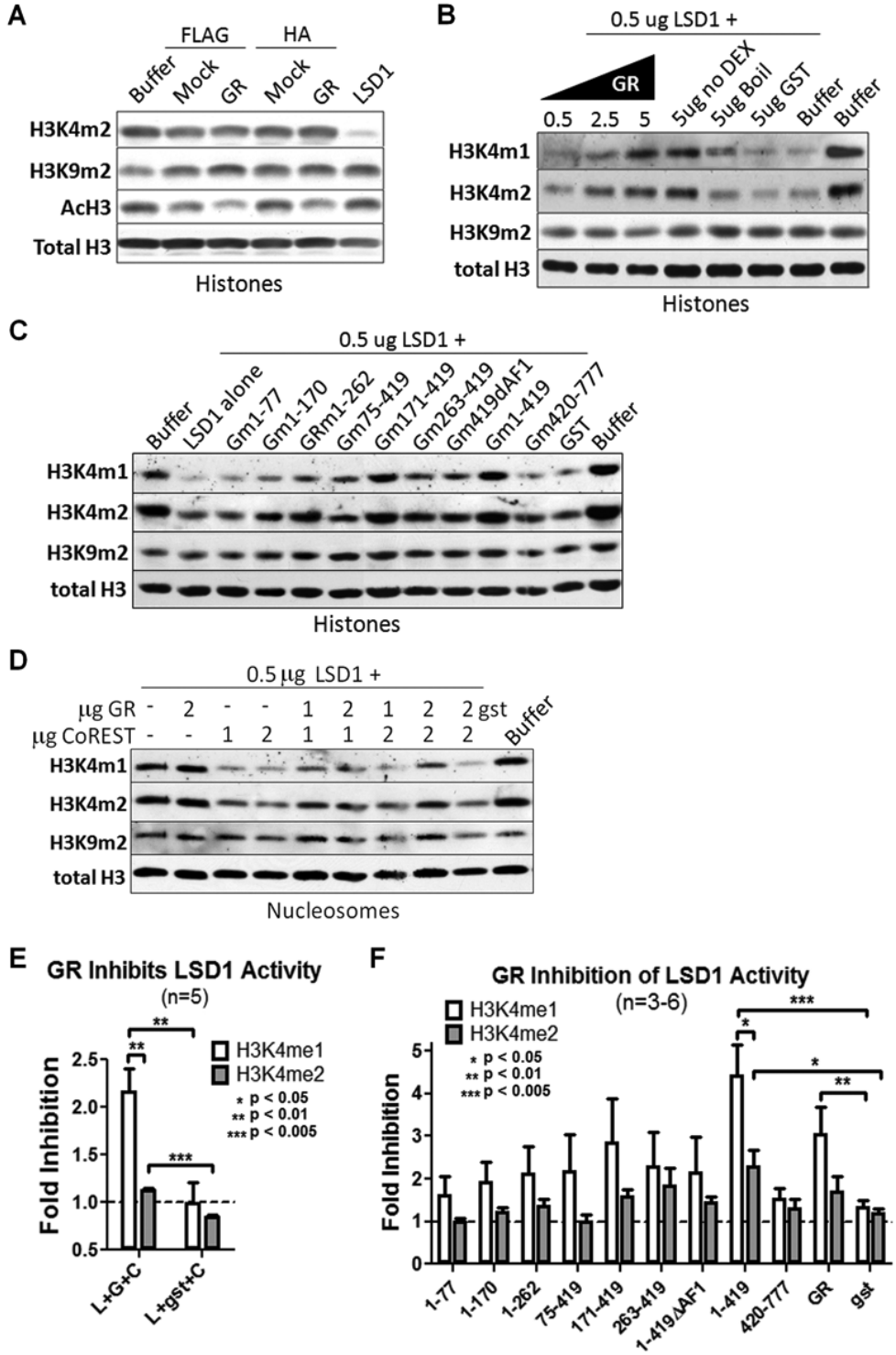


Figure 2.19 GR inhibits LSD1 H3K4, in particular H3K4me1, demethylase activity *in vitro*.

Figure 2.19 (Continued)

Modification specific antibodies were used for Western Blotting to detect enzymatic activity in various assays. Testing the GR TAP complex showed no detectible HDM activity against H3K4 or H3K9, but deacetylation of histones was detected (A). Purified recombinant GR added to LSD1 demethylases of histones inhibited demethylation of H3K4me1/2 (B). GR truncations (Figure 2.3 A) containing the activation also inhibited LSD1 demethylase activity (C) and inhibition of LSD1 was not limited to histones but also occurred when nucleosomes and coREST were used (D). No demethylation of H3K9me2 was detected. Densitometry of demethylase assay Western Blots showed that GR and the GR AD inhibit demethylation of H3K4me1 preferentially over H3K4me2 nucleosomes (E) and histones (F). The average of 3-6 biological replicates is plotted ± standard error, p-values from two-tailed t-test.

NR limiting of LSD1 H3K4 demethylase activity is likely a general phenomenon and novel epigenetic mechanism underlying NR-mediated gene regulation.

Given that LSD1 plays a role in both ER and AR target gene activation, it is possible that regulation of LSD1 H3K4 demethylase activity is an aspect of LSD1 coactivator function for NRs generally. To determine whether or not LSD1 inhibition is unique to GR, several other nuclear receptors were tested in LSD1 HDM assays. Figure 2.20 B shows western blots of HDM assays for estrogen receptor (ER), androgen receptor (AR), a mineralocorticoid receptor N-terminal truncation (MR 1-110aa), thyroid receptor (TR), retinoic acid receptor (RAR), retinoic X receptor (RXR), and RAR-related orphan receptor β (RORβ). All proteins tested except RORβ showed some inhibition of LSD1 activity. Interestingly RORβ lacks a classic N-terminal activation domain (Figure 2.20 A), which may explain the lack of inhibition. These data support a conserved function for NR proteins and specifically the N-terminal activation domain of NRs in inhibiting the H3K4 demethylase activity of LSD1.

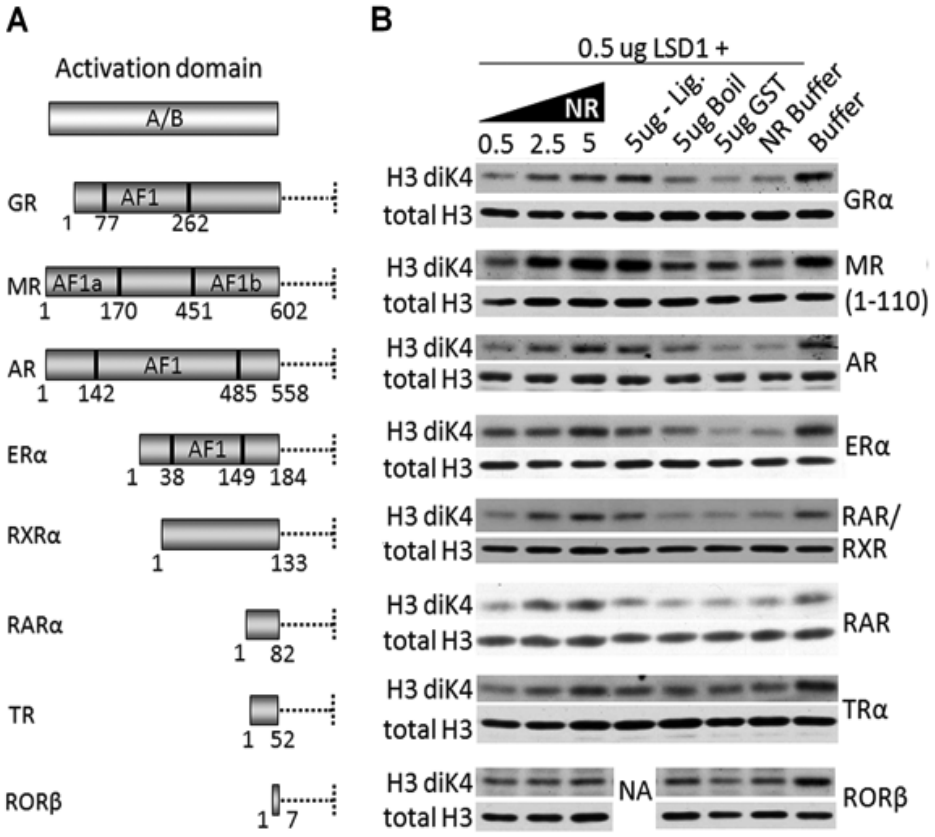


Figure 2.20 Many nuclear receptors inhibit LSD1 H3K4 demethylase activity *in vitro*.

A diagram of the activation domains of NRs tests shows wide variety in structures (A). Multiple NRs inhibit LSD1 demethylase activity when added to *in vitro* histone demethylase assays (B). Inhibition correlates with the presence of an activation domain.

From this we conclude that GR, and more specifically the AD, regulates LSD1 activity by preferentially inhibiting LSD1 H3K4me1 demethylation *in vitro.* This is consistent with our *in vivo* finding that regions bound by both LSD1 and GR show loss of H3K4me2 but accumulation of H3K4me1, while regions without GR (TSSs) show loss of H3K4me2 but no

H3K4me1 accumulation. From this we hypothesize that H3K4me2 loss and H3K4me1 accumulation *in vivo* are dependent on GR-regulated LSD1 activity.

LSD1 is required for changes in H3K4me1/2 and enhancer activation.

To test the above hypothesis, we first examine whether LSD1 activity is responsible for H3K4me2 loss and selective H3K4me1 increases using LSD1 KD of TCP and ChIP-qPCR profiling of H3K4me1/2/3 on the previously defined genes. Neither LSD1 KD nor inhibition of HDM activity prevented GR binding suggesting GR recruitment is independent of LSD1 (Figure 2.21 A) and consistent with RT-qPCR data showing impaired by not fully suppressed activation (Figure 2.9 C).

Also consistent is the finding that H3K4me3 trended toward higher levels and remained restricted primarily to the TSS after DEX treatment similar to controls (Figure 2.21 B), suggesting the addition of this mark is also LSD1-independent. However, changes in H3K4me2/1 and H3K27ac were dependent on LSD1 and its activity. After LSD1 KD no change was observed in H3K4me2 at the TSS or GB enhancers and interestingly, TCP treatment caused increased H3K4me2 levels at both GB enhancers and the TSS (Figure 2.22 A),suggesting LSD1 activity is critical to keep this mark low after DEX-induction. H3K4me1 levels failed to increase at GB enhancers in both KD and TCP cells, while the TSS remained low and unchanged (Figure 2.22 B). Similar to H3K4me1, H3K27ac also failed to increase at GB enhancers suggesting loss of LSD1 or inhibition of its activity disrupts enhancer activation (Figure 2.23).

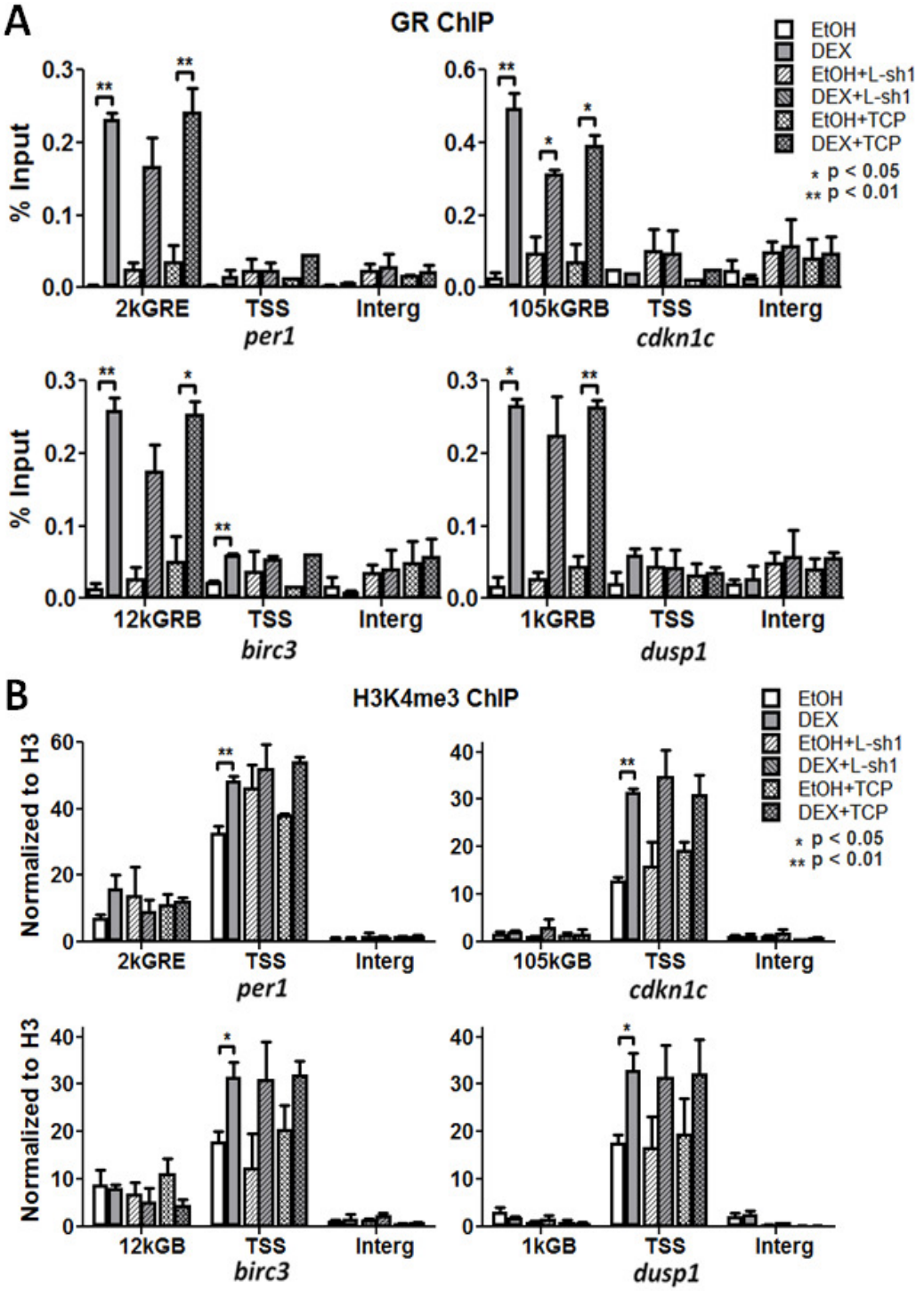


Figure 2.21 GR chromatin binding and H3K4me3 is not disrupted by loss of LSD1 activity.

Figure 2.21 (Continued)
ChIP-qPCR against GR from A549 cells treated with either LSD1 shRNA (L-sh1) or TCP showed DEX-induced GR recruitment independent of LSD1 (A). ChIP-qPCR against H3K4me3 showed trends toward increased levels after DEX treatment (B). Average of 2-3 biological replicates is plotted ± standard error, p-values from two-tailed t-test.

H3K9me2 levels were also examined and found to be disrupted after LSD1 KD or inhibition by TCP. The two sites that showed loss of H3K9me2 in response to DEX no longer decreased and several sites showed an increase in H3K9me2 levels (Figure 2.24) similar to what was observed for H3K4me2 (Figure 2.22 A). This confirms previous findings that LSD1 is required for loss of H3K9me2 in response to hormone-induced gene activation. However, it also supports also supports a possible interdependence between H3K4me2 and H3K9me2.

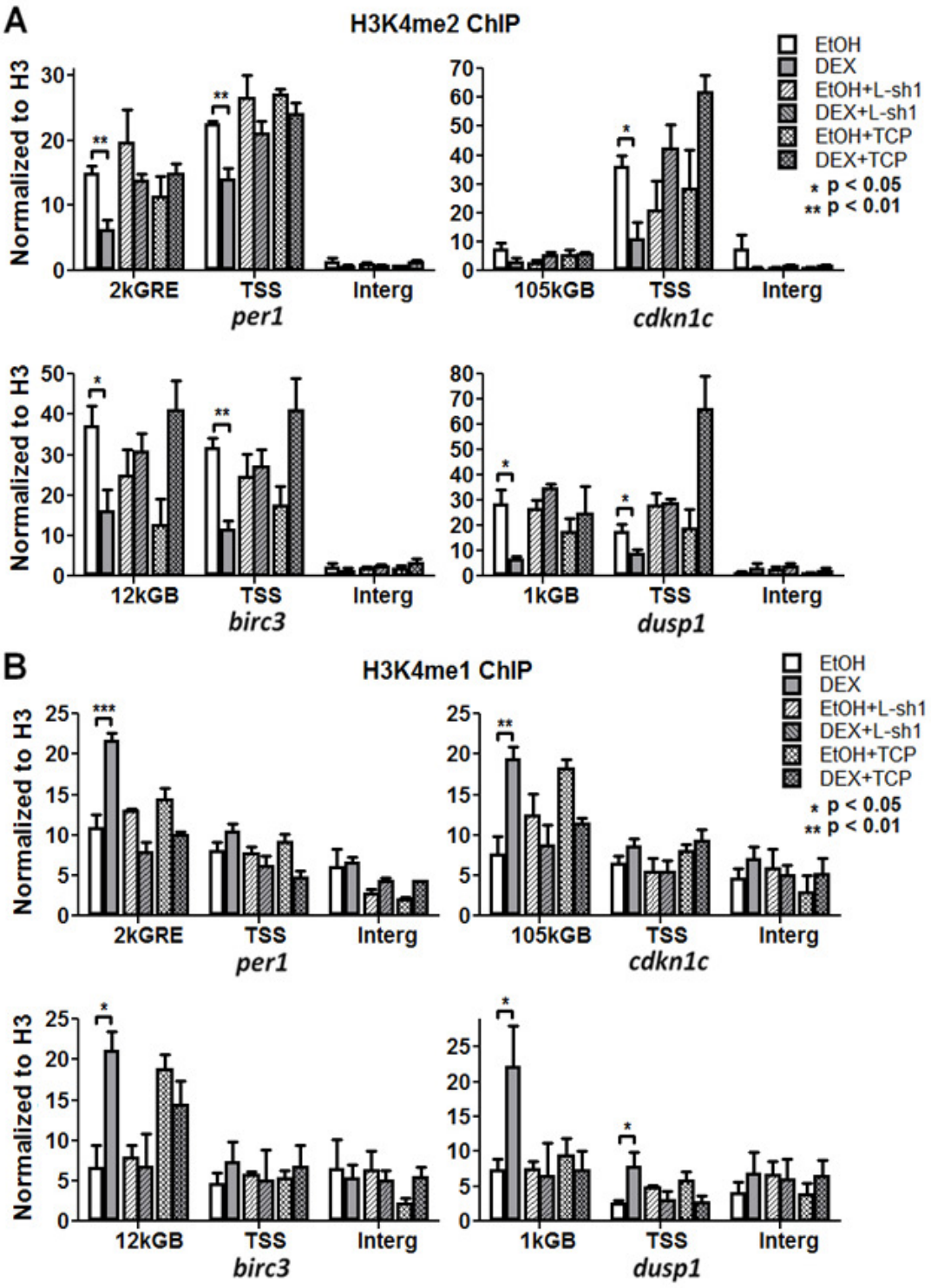


Figure 2.22 LSD1 is a key factor responsible for direct changes in H3K4me1/2.

Figure 2.22 (Continued)

ChIP-qPCR against H3K4me2 (A) or H3K4me1 (B) from A549 cells treated with either LSD1 shRNA (L-sh1) or TCP no longer showed DEX-induced decreases in H3K4me2 or increases in H3K4me1. Average of 2-4 biological replicates is plotted ± standard error, p-values from two-tailed t-test.

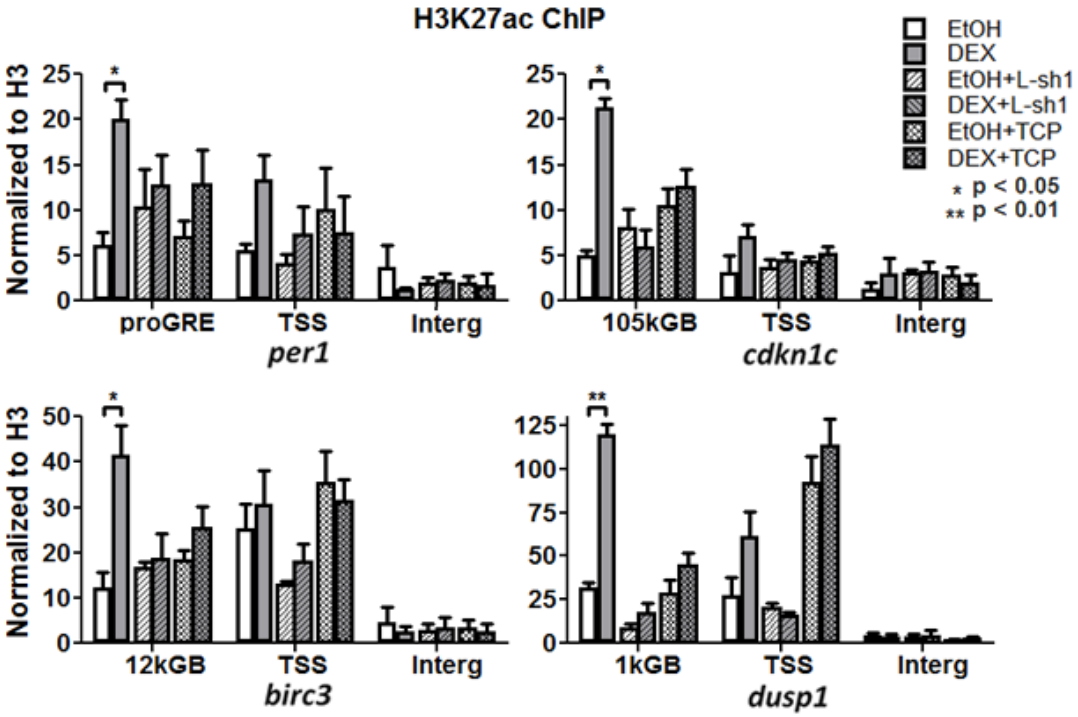


Figure 2.23 LSD1 is a key factor responsible for enhancer activation.

ChIP-qPCR against H3K27ac from A549 cells treated with either LSD1 shRNA (L-sh1) or TCP no longer showed increased levels of the active enhancer mark, H3K27ac, after DEX treatment. Average of 2-4 biological replicates is plotted ± standard error, p-values from two-tailed t-test.

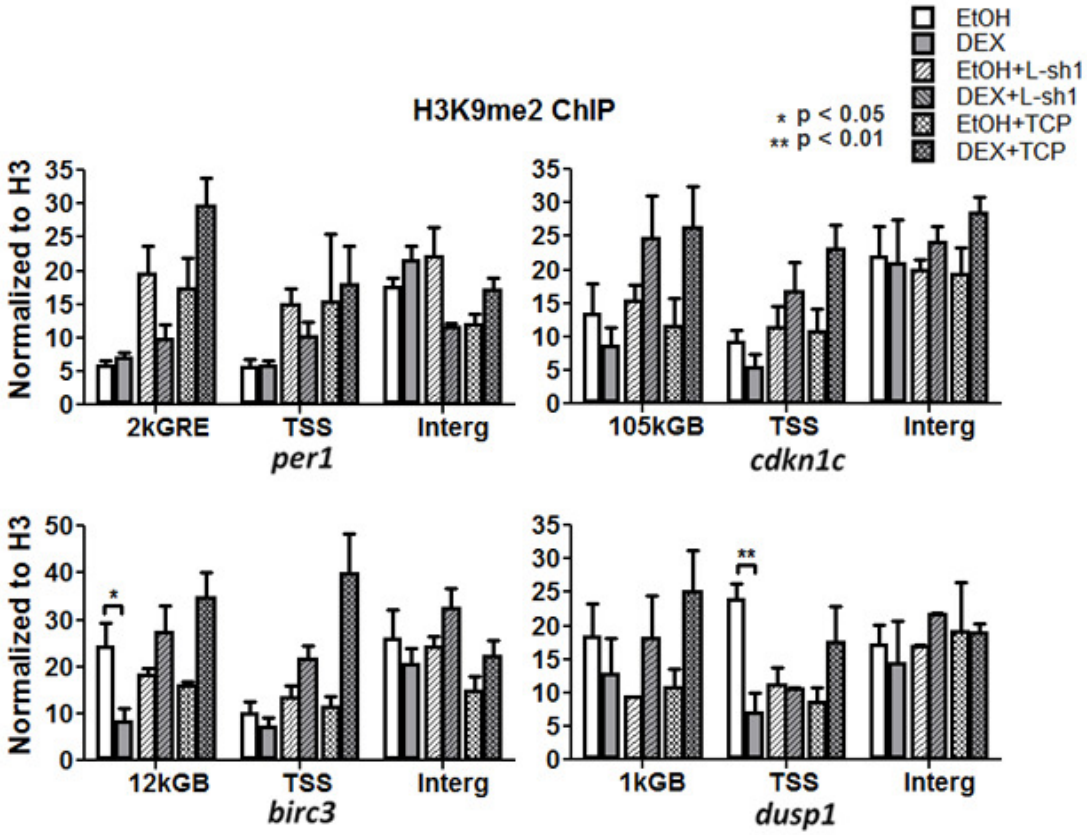


Figure 2.24 LSD1 is a key factor responsible for enhancer activation.
ChIP-qPCR against H3K9me2 from A549 cells treated with either LSD1 shRNA (L-sh1) or TCP no longer showed decreased levels of H3K9me2 after DEX treatment, and a general trend toward increased levels were seen at several other sites. Average of 2-5 biological replicates is plotted ± standard error, p-values from two-tailed t-test.

Discussion

Previous work has showed that LSD1 is an important coactivator for AR and ER gene targets [25-30]. We show here that LSD1 is also an important coactivator for GR, suggesting it may be a general NR coactivator. Combining all of these mechanistic elements we envision

LSD1's role in GR target gene regulation as follows. In the absence of hormone LSD1 is bound to a certain subset of non-specific GREs and possibly removing H3K4me2/1 to mask these sites from future GR activation. Upon hormone exposure and GR translocation to the nucleus LSD1 and GR interact. The methods employed here do not allow us to precisely delineate the timing and location of LSD1 and GR first interaction. We find modest levels of LSD1 at GR bound sites prior to GR binding as well as moderate increases in LSD1 levels upon GR binding. From this data it is impossible to know whether GR can interact with and regulate LSD1 already bound to chromatin or if GR forms a complex with LSD1 prior to chromatin binding and evicts previously bound LSD1 exchanging it for new LSD1 containing complexes. Given the highly dynamic nature of GR DNA binding, it is likely that the true mechanism is somewhere in between these two discrete mechanisms. LSD1 association with DNA may also be highly dynamic and frequent flux between bound and unbound states may provide opportunities for interaction with GR and quick re-association with chromatin.

The role of LSD1 at functional GREs, albeit at low levels, prior to GR binding remains mysterious given our finding of high levels of H3K4me2/1 at these sites. Perhaps this is indicative of LSD1 recruitment to GREs generally, and the presence of other factors determines the functionality of the GRE by regulating LSD1 activity to ultimately mask or tag the GRE for activation. GR is not alone in its ability to regulate LSD1 activity. CoREST facilitates LSD1-mediated demethylation of nucleosomes while BHC80 has been found to inhibit LSD1 activity *in vitro* [24].

LSD1 has been identified in many complexes [20,57,72-78], suggesting that it may be a widely utilized epigenetic tool distributed throughout the genome for various purposes. Perhaps, the key regulatory point for LSD1 is not localization, but instead through specific binding partners with varying abilities to regulate LSD1 activity. For example, unencumbered LSD1 bound to a nonfunctional GRE would remove H3K4me1/2 while LSD1 bound to a functional GRE in complex with a full inhibitor would allow H3K4me1/2 accumulation prior to GR activation. Swapping of the full inhibitor for GR would then allow removal of H3K4me2 by LSD1 and further accumulation of H3K4me1 through conversion from H3K4me2. Swapping sequence specific regulators such as GR while minimizing the changes required of chromatin the tool-kit may provide an advantage for highly dynamic systems such as hormone responses.

Regardless of the exact timing/location, association of LSD1 and GR at a chromatin site results in partial inhibition of LSD1 activity converting the early poised enhancer H3K4me2 mark to H3K4me1 and facilitating further enhancer activation likely through recruitment of other proteins, including an H3K27 acetyltransferase such as CBP/P300. Full enhancer activation then leads to chromatin looping bringing transcription factor complexes in proximity of the promoter and TSS recruiting basal transcription machinery and turning the gene on. Thus LSD1's role in GR target gene regulation occurs at the early stages of target discrimination and enhancer activation.

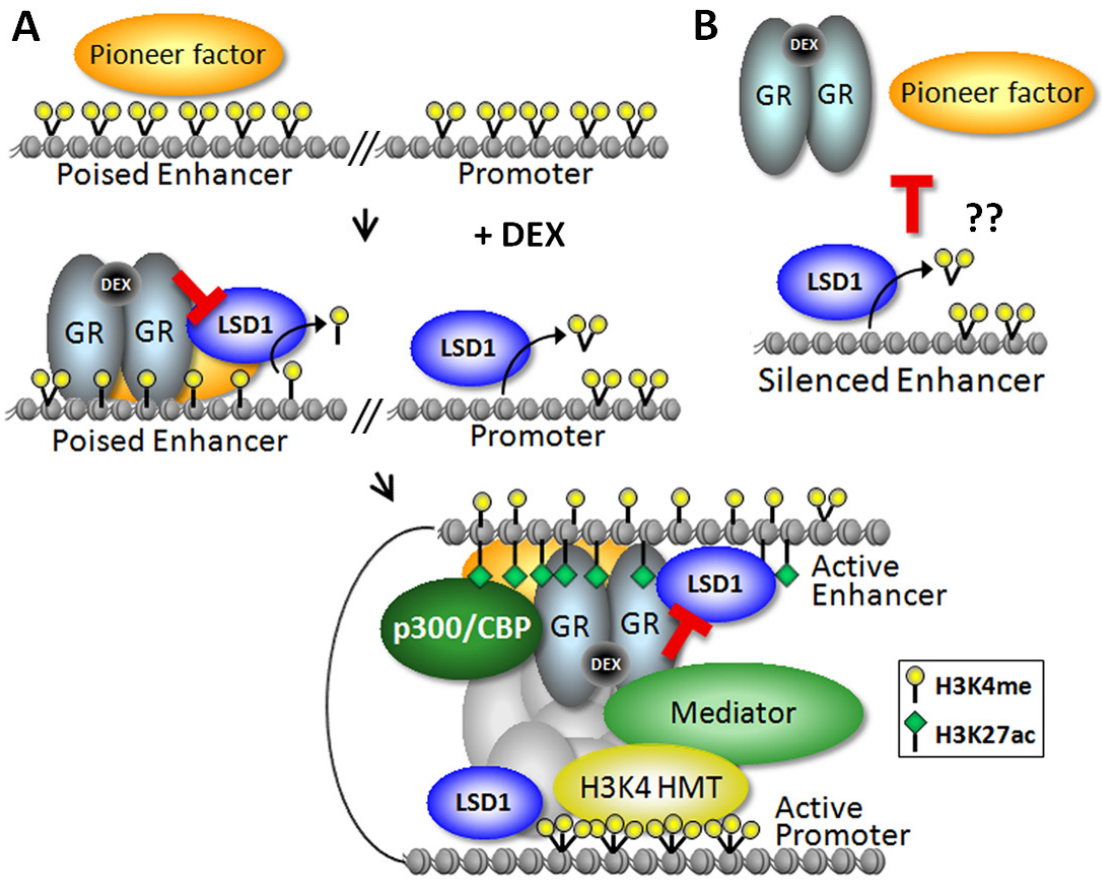


Figure 2.25 Model of LSD1's role in activation of H3K4me2 marked enhancers.

Pioneer factor(s) may be responsible for recognizing and/or protecting specific enhancer sites prior to GC exposure and helping recruit GR after hormone exposure similar to findings with FOXA1 and ER. Once GR and LSD1 are co-localized the H3K4me2 enhancer signature is converted to H3K4me1 by partially inhibited LSD1 and the enhancer is activated with additional modifications such H3K27ac likely laid down by HATs such as p300 found in the GR TAP complex. At promoters recruited LSD1 in the absences of GR removes H3K4me2 and me1. The recruitment of H3K4 HMTs such as MLL2 found in the GR TAP complex counterbalance LSD1 activity to generate a region around the TSS of high H3K4me3 (protected from LSD1 activity) but lacking intermediate states of methylation (A). Microarray data implicates LSD1 silencing non-specific GR targets, and we propose a possible mechanism where LSD1 removes the H3K4me2 enhancer signature helping pre-

Figure 2.25 (Continued)
establish a specific H3K4me2 profile throughout the genome that will allow for cell-type specific GR targeting (B).

Importantly, our work introduces two new epigenetic mechanisms involving LSD1 to the model of GC target gene regulation (Figure 2.25). First is the role of LSD1 in establishing an active enhancer histone code for GC inducible genes as well as maintaining low H3K4me1/2 at the TSS after activation. We propose that GR recruitment and H3K4me3 methylation are independent or prior to LSD1 function, whereas H3K4me1 accumulation is the result of converting H3K4me2 to me1 via precise control of LSD1 activity by GR. To uncover the molecular and biochemical mechanism underlying GR regulation of LSD1 we demonstrate a direct protein-protein interaction between GR and LSD1 and identify the activation domain of GR (1-419aa) as the critical region for interacting and regulating LSD1 *in vitro*. Importantly, we show that direct interaction with GR preferentially inhibits LSD1 H3K4me1 demethylase activity. To our knowledge, this study is the first to report a molecular mechanism by which LSD1 acting as a H3K4 HDM participates in gene activation.

LSD1-dependent enrichment of H3K4me1 and H3K27ac at distal GBs strongly supports epigenetic regulation at enhancers by GR/LSD1 complexes. Hormone-inducible enhancers/genes are thought to be in a type of poised state prior to ligand activation [55,79]. H3K4me2 has been shown to be critical for binding of the NR recruiting, pioneer factor, FoxA1 [39], suggesting this mark may be important for early recognition of functional NR binding sites perhaps in an early poised state. Interestingly, loss of FoxA1 results in

reduced H3K4me1 at many enhancers [80], which supports the role of H3K4me1 in enhancer function but also supports a role for H3K4me2 and pioneer factors upstream of H3K4me1. The present study shows that H3K4me2 was more highly enriched than H3K4me1 at functional GBs prior to GC stimulation. The observed correlation between

LSD1 binding and H3K4me2, as well as the LSD1-dependent drop in H3K4me2 after DEX suggests LSD1 is targeted to its substrate, which is acting as an early mark of inducible enhancers (Figure 2.25A). While studies in ES cells show that H3K4me1 is a mark of poised enhancers that transition to an active state with H3K27ac and p300 binding [38], our study does not distinguish an H3K4me1-only state from one with both H3K4me1 and H3K27ac. This may be due to the rapid nature of DEX induction, unlike the stepwise process of cellular differentiation. We propose the H3K4me2, more so than H3K4me1, represents an early poised state, and/or a poised state unique to rapidly inducible gene systems and that high levels of H3K4me2 precede changes in H3K4me1 and activation. We find that cooperative control of the swift transition from H3K4me2 to H3K4me1 by GR/LSD1 complexes is a prerequisite to initiate enhancer activation.

While we find that H3K27ac is dependent on LSD1 activity (Figure 2.23) it is unlikely that LSD1 is directly responsible, but instead that this mark is dependent on H3K4me1 accumulation. This is consistent with the characterization of H3K27ac as a mark that follows H3K4me1 during ESC differentiation [38]. However, given the associated loss of H3K4me2, the simplest explanation for H3K4me1 accumulation ***is*** as a direct result of LSD1-mediated removal of one methyl group. The findings that H3K4me1 only accumulates

on LSD1 bound sites also bound by GR, and that GR inhibits LSD1-mediated demethylation of H3K4me1 *in vitro* support a mechanism where GR partially inhibits LSD1 activity allowing conversion of H3K4me2 to H3K4me1. Another possible explanation for H3K4me1 accumulation not addressed by this study is the recruitment of an H3K4 methyltransferase to GB enhancers. In Drosophila MLL3/4 was shown to specifically mono-methylate H3K4 at enhancers and MLL2 was identified in our GR complex (Figure 2.2 C). While it is possible that an MLL complex recruited by GR to enhancers is also involved in H3K4me1 accumulation, the dependence of this mark on LSD1 would suggest at the very least cooperation between HMT and HDM activities. However, in such a mechanism we would predict that loss of LSD1 would disrupt the balance between HMT/HDM driving H3K4me2/3 levels up at GB enhancers, which was not observed after LSD1 KD (Figure 2.21 B and 2.22 A). Taking these data together we propose a molecular model where LSD1 is recruited to sites with high H3K4me2, and at sites also bound by GR, LSD1 is partially inhibited converting some H3K4me2 to H3K4me1 and thus promoting H3K27 acetylation and enhancer activation.

Induction of LSD1 binding at the TSS of DEX-regulated genes and the loss of H3K4me2 at these sites suggest a contribution to the mechanism of gene activation. Although H3K4me2 is generally associated with activation and thus its loss should contribute to silencing, its role may be more nuanced. In yeast, H3K4me2 has been implicated as a type of memory mark of recent transcription at the promoters of inducible genes, which needs to be removed for full reactivation [81]. This mechanism may relate to the methods employed in this study, where cells were grown in normal serum containing many hormones prior to

growth in stripped serum for 16-24 hours and then reintroduction of an activating signal by DEX. The high levels of H3K4me2 at the TSS may be a type of memory mark of previous expression in normal serum. A second explanation might involve conversion of H3K4me2 to me3, likely due to recruitment of an HMT complex such as MLL2 to the TSS. Again LSD1 may play a role in counterbalancing HMT activity at the TSS to generate a region that is primarily trimethylated. This is supported by the observation that H3K4me2 levels increase along with H3K4me3 after TCP treatment suggesting LSD1 helps maintain low levels of H3K4me2 while H3K4me3 accumulates. While HMT and HDM activities are typically thought of as functionally counteracting one another, i.e. K4 demethylation as repressive and K4 methylation as activating, these opposing reactions could also be combined to generate highly specific methyl states. For example, LSD1 can only remove H3K4me1/2 but not me3. Therefore, co-localization of LSD1 with a HMT could result in low levels of H3K4me1/2 but also high levels of H3K4me3 as this mark is protected from LSD1. Pairing HDMs and HMTs with different specificities and/or regulating the enzymatic activities through protein-protein interactions could fine-tune the overall reaction to favor a single methyl state.

Interestingly, loss of H3K4me2 at E2 induced promoters regulated by LSD1 and ER has been previously observed [30]. However, this and other studies focused on loss of H3K9me1/2 at hormone-inducible promoters and with good reason given that H3K9 methylation is a well-established repressive mark. The finding that loss of H3K9me1/2 is LSD1-dependent has lead to the conclusion that LSD1 acts as an H3K9me1/2 demethylase directly. We also find LSD1-dependent loss of H3K9me2 at a small number of GR bound

sites and GC regulated genes. However, given that only a subset of sites showed changes in H3K9me2 while most showed changes in H3K4me2 we conclude H3K9me2 likely plays a minor or perhaps highly specialized role in regulating GR target genes. We also propose that the LSD1 dependence of H3K9me2 demethylation could be due to interdependence of histone marks, an aspect of the histone code that is very poorly understood in general, let alone in NR-mediated transcriptional regulation. Demonstrating direct enzymatic activity *in vivo* is a significant challenge in enzymatic characterizations, thus *in vitro* evidence is a mainstay. Contrary to findings with ER and AR, our *in vitro* characterization of LSD1 enzymatic activity in the GR complex found no evidence of H3K9 demethylation suggesting this activity may be carried out by another HDM. In the case of AR it was proposed that JMJD2C, an H3K9m2/3 and H3K36me2/3 demethylase, and LSD1 cooperated to demethylate H3K9 [27]. In the GR TAP complex we find JMJD1C, and likely H3K9me1/2 demethylase that has also interacts with AR [82,83]. While we do not rule out a role for LSD1 in H3K9 loss *in vivo* or a contribution by this demethylation event in GC-induced gene activation, we do suggest the mechanistic details are more complex than previously described. The finding that changes in H3K4 methylation, a well-established LSD1 substrate, are part of the activation mechanism and dependent on LSD1 suggests this may be LSD1's primary role. Also, the observation that both H3K9me2 and H3K4me2 levels track similar increases after TCP treatment (Figures 2.22 A and 2.24) warrants a closer look at potential relationships between H3K4 and H3K9 methylation.

The second role for LSD1 in GR target gene regulation is in suppressing non-specific targets. Expression profiling of DEX responsive genes after LSD1 KD indicates a role for

LSD1 in suppressing non-specific targets (Groups E and F, Figure 2.8 B). This combined with our finding that H3K4me2 is enriched at functional GR binding sites and LSD1's ability to remove H3K4 methylation implicate it in regulating a cell-type specific early poised enhancer code prior to hormone exposure. One plausible mechanism could be that in the absence of GR, LSD1 removes the H3K4me2 (as well as H3K4me1) signature from non-specific enhancers (Figure 2.25 B), while functional sites are protected by some mechanism that preserves the H3K4me2 signal. Upon GC exposure and GR entering the nucleus it binds only those predetermined sites allowing for rapid as well as cell-type specific responses.

This mechanism is reminiscent the original characterization of LSD1 function in the REST complex, where LSD1 is responsible for repressing neuronal genes in non-neuronal cell types [71,84]. Similar to the REST complex, LSD1's role in suppressing non-specific GR targets likely involves multiple other protein regulators responsible for the specificity of distinguishing which sites should and shouldn't retain H3K4me2. It would be interesting for future studies to identify the HMT and other regulatory factors responsible for laying down the H3K4me2 signature and distinguishing functional from non-functional GR binding sites prior to stimulation. Understanding regulation at this crucial step could lead to important insights into the mechanisms of cell-type specific hormone action.

Our understanding of the histone code often focuses on the offensive side of the equation as we translate the meanings of various histone marks. However, for every histone writer we now understand there is also likely a corollary histone eraser and the balance and targeting

of these two activities can almost certainly be tweaked in ways and degrees we don't yet fully understand.

While LSD1 KD impacts a majority of DEX-responsive genes, it is interesting to note that many were unaffected. Decreasing the stringency for defective activation or repression from 50% to 30% still only brings the number of affected genes to 73%, suggesting there may be mechanistically divergent subsets of GC regulated genes, some relying on LSD1 and others that do not. The wide range in H3K4me2 prior to GR binding (Figure 2.18 A) suggests this mark may also contribute more to the recognition mechanism of some sites than others. Dissecting GC-responsiveness into LSD1-dependent and -independent mechanisms could have important translational implications. The potent anti-inflammatory actions of GCs are limited by significant side effects such as osteoporosis and metabolic disruptions. This has lead researchers to look for selective receptor modulators (SRMs) that can dissociate positive and negative effects [85]. However, a molecule with the ability to satisfactorily separate the good from the bad has yet to be discovered. GO analysis of DEX-induced genes showed several known DEX-regulated pathways, such as pro-survival and inflammation. While pro-survival was also a top LSD1-dependent pathway along with proliferation, inflammation dropped out suggesting it was not heavily influenced by LSD1 coactivator function. The GO analysis was supported by cellular assays for DEX and LSD1 affects on apoptosis, proliferation, and IL8 release. While LSD1 was important for DEX-regulation of apoptosis and proliferation, it was not important for DEX-suppression of IL8 mRNA levels or TNFα induced IL8 secretion (Figure 2.13 A and B). While IL8 is not the only point of GC immune regulation and experiments in cell culture do

not always predict outcomes at the organism level, our data does support further investigation into the possibility of modulating GC physiological actions through coregulators such as LSD1.

Similarly, the new mechanisms illustrated in this study could also have significant therapeutic implications in the treatment of cancer. The role of GCs in cell proliferation and survival is highly cell type specific. GCs block cell division in lymphoid, fibroblastic, and epithelial tissues [86-88], and induce apoptosis in lymphoid cell lineages such as T-cells and monocytes, contributing to immunosuppression and anti-inflammatory effects, but also their use in treating lymphoid derived cancers [43-45]. However, GCs protect from apoptosis in several cell types [46-50], making their use as a co-treatment to reduce nausea and alleviate the acute toxic effects of chemotherapy questionable. For example, concurrent GC and chemotherapy treatment has been shown to inhibit chemotherapy-induced apoptosis in Bcap37 breast cancer [89], HL-60 human leukemia [90], human glioma and rhabdomyosarcoma [91], human urothelial [92], and human cervical and lung cancer cell lines [93].

Our findings suggest LSD1 inhibitors could be useful in blocking the pro-survival affects of GCs thus assuaging concerns about co-treatment. Additionally, LSD1 inhibitors could potentially sensitize epithelial derived tumors to chemotherapeutic induced apoptosis, adding value to the practice of using GCs in chemotherapy regimes. On the whole we think the nature of LSD1's role in GR-mediated gene regulation and the importance of GC action both in hormonal regulation of human physiology as well as clinical applications make it a critical coregulator to understand and an appealing therapeutic target.

Experimental Procedures

Cell culture

HeLa-s and A549 cells were cultured in DMEM supplemented with 20% FBS. MDA Kb2 cells (ATCC: CRL-2713) were cultured in RPMI 1640 supplemented with 20% FBS. All DEX treatments were carried out after cells were grown in media supplemented with charcoal striped serum for 16-24 hours.

Proteins, Recombinant protein expression, purification, and GST-pulldown

Purified TR, ER, MR (1-110aa), ROR were purchased from Abnova (catalog ID: H00007067-P01, H00002100-P01, H00004306-Q01, H00006096-P01) and AR was purchased from RayBiotech (RB-14-0003P). GST-RAR and RXR was a gift from Dr. Anders Näär. GR and GR truncation mutants were expressed and purified from the pGEX-4T-1 (27-458001) vector in E. coli. The homogeneity of the eluted protein was determined using SDS–PAGE followed by Coomassie blue staining.

For GST pulldown, GR proteins were incubated with 6xHIS-LSD1 overnight at 4°C in buffer containing 200mM NaCl and 0.1% NP-40.

Primers and Antibodies

ChIP antibodies were, LSD1 (Abcam ab17721), GR-H300 (Santa Cruz sc-8992), H3K4me1 (Abcam ab8895), H3K4me2 (Millipore 07-030), H3K4me3 (Millipore 07-473), H3K9me2 (Abcam ab1220), H3K27ac (Millipore 17-683), and H3 (Abcam ab1791). The amplification

efficiency and specificity of each set of primers was obtained by standard curve analysis and PCR products were separated by 2% agarose gel electrophoresis visualized by ethidium bromide staining.

ChIP-qPCR primer sets

BIRC3_TSS	GGTTATTACCGCTGGAGTTC AAATGCGTCACCCAAATCC
BIRC3_GRB	GATGGCCAGTAATGGAACTG ATGCATCTCATCAGGGCATC
CDKN1C _TSS	ACTAGTACTGGGAAGGTCC TTCTTCTCGCTGTCCTCTC
CDKN1C _GRB	AGGTCAGCTCACAGGATTG CCCTTGCGCAAAGAGAAAG
DUSP1_TSS	GTCAGACCACTTAACTGTGG GCAAAGGCATGGAAGAGTAG
DUSP1_GRB	CCAGGTGCATTACAGGTATC CTTAGGCATGTGACCTTTGG
PER1_TSS	CATCATGTTCTCTTGGCTGGTGG AGGACGGCTGTCGTTTTGTTG
PER1_GRE	CATCAGATTGGAAGTGGCAG CGACCAGGTAGGCATCTC

RNA interference, TCP treatments, and RT–PCR

Retroviral shRNA (Sigma) targeting human LSD1 (L-sh1:TRCN0000046071, L-sh2:TRCN0000046072) and control scramble (SC) shRNA was used to infect MDA-Kb2 and A549 cells. The knockdown efficiency was determined by RT–PCR and western blot. For all shRNA KD experiments cells were grown in selection media for 5-7 days prior to harvesting. For all TCP experiments cells were also treated for 5-7 days. RNA was purified by Trizol (Life Science) and cDNA produced using SuperScriptIII RT kit (Invitrogen)

according to manufacturer's instructions. Sequences of quantitative RT-PCR primers are available upon request.

Complex purification and

GR-associated protein complexes were purified from HeLa-S cells stably expressing FLAG-HA-GR by sequential anti-FLAG and anti-HA Affinity Gel immunoprecipitation as previously described [94]. Cells were treated with 50nM DEX for 6 hours. Associated proteins were sequenced by MS/MS at the Harvard Medical School Taplin Biological Mass Spectrometry Facility. Half of the single FLAG purification or the doubly purified GR complexes, were used for another purification by anti-LSD1 and the associated peptide.

Microarray

A549 cells were infected with retrovirus carrying LSD1 shRNA and selected for 5 days in puromycin prior to treatment with 100nM DEX for 12 hours followed by Trizol (Life Science) RNA extraction. mRNA levels were profiled by Affimetrix 2.0 Microarray chip.

ChIP and ChIP-Sequencing

Conventional ChIP was performed as previously described [94] using formaldehyde-crosslinked chromatin with modification to some buffers detailed in supplemental procedures. Briefly, cells were treated with 100nM DEX or ethanol for 2 hours and then cross-linked with 1% formaldehyde for 10 min at room temperature. Cross-linking was stopped by the addition of glycine to 0.125 M. Cells were washed in cold PBS, suspended in buffer (50mM Tris-HCl pH 7.9, 500mM NaCl, 1mM EDTA, 1% Triton-X100, 0.1% Na-

deoxylcholate) and sonicated to obtain DNA fragments with an average size of 500 base pairs. Sonicated chromatin was diluted in half and pre-cleared using IgG, protein A and G beads blocked with BSA/salmon sperm. Incubation with target antibody was carried out overnight at 4°C. Samples were washed twice with sonication buffer, low salt buffer (20mM Tris-HCl pH 8.0, 150mM NaCl, 2mM EDTA, 0.1% SDS, and 1% Triton X-100), high-salt buffer (same as low but with 500mM NaCl) and TE. Eluted, decrosslinked DNA was purified by phenol:chloroform extraction and enrichment of target regions was determined by qPCR (Bio-Rad IQ5) using SYBR Green reagent (Bio-Rad).ChIP-Seq procedure was similar except sonicated fragments were reduced to an average size of 300bp.

ChIP-Seq was carried out following a similar protocol. Purified DNA was used to generate libraries and sequenced by the Illumina Genome Analyzer II (GA II) per manufacturer's instructions up to 36 cycles. Image analysis and base calling were performed with the Illumina package OLB (v1.8). Sequence reads were mapped onto the reference human genome (NCBI Build UCSC hg19) using the Bowtie (v0.12.7) algorithm. Using the Model based Analysis of ChIP-Seq (MACS) package [95] regions of significant enrichment were determined against sample input. Refseq gene annotation was obtained from the UCSC website. GR peaks were identified using MACS (v1.4) at P<1e-5. Distributions were drawn using in-house software and normalized density. Quantification of LSD1 and H3K4 methyl mark densities for bar graph and ANOVA comparisons were carried out by a density counting program written in-house. Briefly the program generates a density value for a specific DNA region (e.g. the region spanning a GR peak) by calculating the tag enrichment

area taking into account signal height and width. Heatmaps were generated using a 100bp window through GR peaks centered ± 2000bp.

CCK8 proliferation, TUNEL apoptosis, and IL-8 ELISA assays

For the CCK8 assays (Dojindo) A549 cells were treated with shRNA against LSD1 (L-sh1 or L-sh2), GR (GR-sh), or LSD2 (L2-sh) and selected for 3 days prior to treatment with either EtOH or various concentrations of DEX for an additional 5 days. Proliferation was detected by CCK8 activity using the BLANK kit according to manufacturer's instructions.

For TUNEL assays (Roche) A549 cells were treated with shRNA against LSD1 (L-sh1) and selected for 5-7 days prior to fixation and staining according to manufacturer's instructions. Cells were analyzed by flow cytometry.

For IL8 ELISA assays (Biolegend) A549 cells were treated with shRNA against LSD1 (L-sh1) and selected for 3-4 days prior to plating into 96-well plates. After allowing cells to attach for 24 hours they were treated with DEX for 6 hours followed by a 24 hour treatment with 2ng/ml TNFα. Cells were then washed with PBS and lysed in RIPA buffer for ELISA analysis according to manufacturer's instructions.

Histone demethylase assay

Histone demethylase assays were carried out as described previously [96]. Briefly, 6xHIS-LSD1 was incubated with purified NRs, GR mutants, and/or CoREST for 1h at 4°C prior to overnight incubation at 32°C with bulk histones or purified nucleosomes in demethylase

buffer (50mm Tris, pH 8.0, 0.1 Units formaldehyde dehydrogenase (FDH) and 1mM NAD+). In a typical reaction, 4µg of calf thymus histones or 6 µg of purified nucleosomes were incubated with 0.5µg LSD1 and 0.5-2.5µg NR in a total reaction volume of 50µl. Reactions were analyzed by SDS–PAGE and western blotting using methyl-specific antibodies.

References

1. Silverman, M.N. & Sternberg, E.M. Glucocorticoid regulation of inflammation and its functional correlates: from HPA axis to glucocorticoid receptor dysfunction. Annals of the New York Academy of Sciences 1261, 55-63 (2012).

2. Barnes, P.J. Anti-inflammatory actions of glucocorticoids: molecular mechanisms. Clinical science 94, 557-72 (1998).

3. Weinberger, C. et al. Identification of human glucocorticoid receptor complementary DNA clones by epitope selection. Science 228, 740-2 (1985).

4. McKay, L.I. & Cidlowski, J.A. Molecular control of immune/inflammatory responses: interactions between nuclear factor-kappa B and steroid receptor-signaling pathways. Endocrine reviews 20, 435-59 (1999).

5. Dostert, A. & Heinzel, T. Negative glucocorticoid receptor response elements and their role in glucocorticoid action. Current pharmaceutical design 10, 2807-16 (2004).

6. Stafford, J.M., Wilkinson, J.C., Beechem, J.M. & Granner, D.K. Accessory factors facilitate the binding of glucocorticoid receptor to the phosphoenolpyruvate carboxykinase gene promoter. The Journal of biological chemistry 276, 39885-91 (2001).

7. Schoneveld, O.J., Gaemers, I.C. & Lamers, W.H. Mechanisms of glucocorticoid signalling. Biochimica et biophysica acta 1680, 114-28 (2004).

8. Kato, S., Yokoyama, A. & Fujiki, R. Nuclear receptor coregulators merge transcriptional coregulation with epigenetic regulation. Trends in biochemical sciences 36, 272-81 (2011).

9. Biddie, S.C., John, S. & Hager, G.L. Genome-wide mechanisms of nuclear receptor action. Trends in endocrinology and metabolism: TEM 21, 3-9 (2010).

10. Kino, T., Nordeen, S.K. & Chrousos, G.P. Conditional modulation of glucocorticoid receptor activities by CREB-binding protein (CBP) and p300. The Journal of steroid biochemistry and molecular biology 70, 15-25 (1999).

11. Yoshinaga, S.K., Peterson, C.L., Herskowitz, I. & Yamamoto, K.R. Roles of SWI1, SWI2, and SWI3 proteins for transcriptional enhancement by steroid receptors. Science 258, 1598-604 (1992).

12. Henriksson, A. et al. Role of the Ada adaptor complex in gene activation by the glucocorticoid receptor. Molecular and cellular biology 17, 3065-73 (1997).

13. McKenna, N.J. et al. Nuclear receptor coactivators: multiple enzymes, multiple complexes, multiple functions. The Journal of steroid biochemistry and molecular biology 69, 3-12 (1999).

14. Freedman, L.P. Increasing the complexity of coactivation in nuclear receptor signaling. Cell 97, 5-8 (1999).

15. McKenna, N.J. & O'Malley, B.W. Minireview: nuclear receptor coactivators--an update. Endocrinology 143, 2461-5 (2002).

16. John, S. et al. Interaction of the glucocorticoid receptor with the chromatin landscape. Molecular cell 29, 611-24 (2008).

17. Johnson, T.A., Elbi, C., Parekh, B.S., Hager, G.L. & John, S. Chromatin remodeling complexes interact dynamically with a glucocorticoid receptor-regulated promoter. Molecular biology of the cell 19, 3308-22 (2008).

18. Ito, K., Barnes, P.J. & Adcock, I.M. Glucocorticoid receptor recruitment of histone deacetylase 2 inhibits interleukin-1beta-induced histone H4 acetylation on lysines 8 and 12. Molecular and cellular biology 20, 6891-903 (2000).

19. Whyte, W.A. et al. Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. Nature 482, 221-5 (2012).

20. Wang, J. et al. Opposing LSD1 complexes function in developmental gene activation and repression programmes. Nature 446, 882-7 (2007).

21. Shao, G.B. et al. [Histone demethylase LSD1 and its biological functions]. Yi chuan = Hereditas / Zhongguo yi chuan xue hui bian ji 32, 331-8 (2010).

22. Chen, Y., Jie, W., Yan, W., Zhou, K. & Xiao, Y. Lysine-specific histone demethylase 1 (LSD1): A potential molecular target for tumor therapy. Critical reviews in eukaryotic gene expression 22, 53-9 (2012).

23. Lee, M.G., Wynder, C., Cooch, N. & Shiekhattar, R. An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. Nature 437, 432-5 (2005).

24. Shi, Y.J. et al. Regulation of LSD1 histone demethylase activity by its associated factors. Molecular cell 19, 857-64 (2005).

25. Nair, S.S. et al. PELP1 is a reader of histone H3 methylation that facilitates oestrogen receptor-alpha target gene activation by regulating lysine demethylase 1 specificity. EMBO reports 11, 438-44 (2010).

26. Metzger, E. et al. Phosphorylation of histone H3T6 by PKCbeta(I) controls demethylation at histone H3K4. Nature 464, 792-6 (2010).

27. Wissmann, M. et al. Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression. Nature cell biology 9, 347-53 (2007).

28. Metzger, E. et al. LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. Nature 437, 436-9 (2005).

29. Perillo, B. et al. DNA oxidation as triggered by H3K9me2 demethylation drives estrogen-induced gene expression. Science 319, 202-6 (2008).

30. Garcia-Bassets, I. et al. Histone methylation-dependent mechanisms impose ligand dependency for gene activation by nuclear receptors. Cell 128, 505-18 (2007).

31. Forneris, F., Binda, C., Battaglioli, E. & Mattevi, A. LSD1: oxidative chemistry for multifaceted functions in chromatin regulation. Trends in biochemical sciences 33, 181-9 (2008).

32. Hou, H. & Yu, H. Structural insights into histone lysine demethylation. Current opinion in structural biology 20, 739-48 (2010).

33. Upadhyay, A.K., Horton, J.R., Zhang, X. & Cheng, X. Coordinated methyl-lysine erasure: structural and functional linkage of a Jumonji demethylase domain and a reader domain. Current opinion in structural biology 21, 750-60 (2011).

34. So, A.Y., Chaivorapol, C., Bolton, E.C., Li, H. & Yamamoto, K.R. Determinants of cell- and gene-specific transcriptional regulation by the glucocorticoid receptor. PLoS genetics 3, e94 (2007).

35. Reddy, T.E. et al. Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. Genome research 19, 2163-71 (2009).

36. Carroll, J.S. et al. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell 122, 33-43 (2005).

37. Tang, Q. et al. A comprehensive view of nuclear receptor cancer cistromes. Cancer research 71, 6940-7 (2011).

38. Creyghton, M.P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences of the United States of America 107, 21931-6 (2010).

39. Lupien, M. et al. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell 132, 958-70 (2008).

40. Heintzman, N.D. et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459, 108-12 (2009).

41. Wang, J.C. et al. Chromatin immunoprecipitation (ChIP) scanning identifies primary glucocorticoid receptor target genes. Proceedings of the National Academy of Sciences of the United States of America 101, 15603-8 (2004).

42. Rogatsky, I. et al. Target-specific utilization of transcriptional regulatory surfaces by the glucocorticoid receptor. Proceedings of the National Academy of Sciences of the United States of America 100, 13845-50 (2003).

43. Compton, M.M. & Cidlowski, J.A. Rapid in vivo effects of glucocorticoids on the integrity of rat lymphocyte genomic deoxyribonucleic acid. Endocrinology 118, 38-45 (1986).

44. Schmidt, M. et al. Glucocorticoids induce apoptosis in human monocytes: potential role of IL-1 beta. Journal of immunology 163, 3484-90 (1999).

45. Thompson, E.B., Webb, M.S., Miller, A.L., Fofanov, Y. & Johnson, B.H. Identification of genes leading to glucocorticoid-induced leukemic cell death. Lipids 39, 821-5 (2004).

46. Evans-Storms, R.B. & Cidlowski, J.A. Delineation of an antiapoptotic action of glucocorticoids in hepatoma cells: the role of nuclear factor-kappaB. Endocrinology 141, 1854-62 (2000).

47. Gorman, A.M., Hirt, U.A., Orrenius, S. & Ceccatelli, S. Dexamethasone pre-treatment interferes with apoptotic death in glioma cells. Neuroscience 96, 417-25 (2000).

48. Chae, H.J. et al. Dexamethasone suppresses tumor necrosis factor-alpha-induced apoptosis in osteoblasts: possible role for ceramide. Endocrinology 141, 2904-13 (2000).

49. Wen, L.P., Madani, K., Fahrni, J.A., Duncan, S.R. & Rosen, G.D. Dexamethasone inhibits lung epithelial cell apoptosis induced by IFN-gamma and Fas. The American journal of physiology 273, L921-9 (1997).

50. Kim, Y.S., Park, J.S., Jee, Y.K. & Lee, K.Y. Dexamethasone inhibits TRAIL- and anti-cancer drugs-induced cell death in A549 cells through inducing NF-kappaB-independent cIAP2 expression. Cancer research and treatment : official journal of Korean Cancer Association 36, 330-7 (2004).

51. Mikosz, C.A., Brickley, D.R., Sharkey, M.S., Moran, T.W. & Conzen, S.D. Glucocorticoid receptor-mediated protection from apoptosis is associated with induction of the serine/threonine survival kinase gene, sgk-1. The Journal of biological chemistry 276, 16649-54 (2001).

52. Krum, S.A. et al. Unique ERalpha cistromes control cell type-specific gene regulation. Molecular endocrinology 22, 2393-406 (2008).

53. Eeckhoute, J., Carroll, J.S., Geistlinger, T.R., Torres-Arzayus, M.I. & Brown, M. A cell-type-specific transcriptional network required for estrogen regulation of cyclin D1 and cell cycle progression in breast cancer. Genes & development 20, 2513-26 (2006).

54. Clarke, C.L. & Graham, J.D. Non-overlapping progesterone receptor cistromes contribute to cell-specific transcriptional outcomes. PloS one 7, e35859 (2012).

55. John, S. et al. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. Nature genetics 43, 264-8 (2011).

56. Saleque, S., Kim, J., Rooke, H.M. & Orkin, S.H. Epigenetic regulation of hematopoietic differentiation by Gfi-1 and Gfi-1b is mediated by the cofactors CoREST and LSD1. Molecular cell 27, 562-72 (2007).

57. Wang, Y. et al. LSD1 is a subunit of the NuRD complex and targets the metastasis programs in breast cancer. Cell 138, 660-72 (2009).

58. Yu, C.Y. et al. Genome-wide analysis of glucocorticoid receptor binding regions in adipocytes reveal gene network involved in triglyceride homeostasis. PLoS One 5, e15188.

59. Phuc Le, P. et al. Glucocorticoid receptor-dependent gene regulatory networks. PLoS Genet 1, e16 (2005).

60. Cheng, X., Zhao, X., Khurana, S., Bruggeman, L.A. & Kao, H.Y. Microarray analyses of glucocorticoid and vitamin D3 target genes in differentiating cultured human podocytes. PLoS One 8, e60213.

61. Kuo, T. et al. Genome-wide analysis of glucocorticoid receptor-binding sites in myotubes identifies gene networks modulating insulin signaling. Proc Natl Acad Sci U S A 109, 11160-5.

62. Gregoire, F.M., Smas, C.M. & Sul, H.S. Understanding adipocyte differentiation. Physiological reviews 78, 783-809 (1998).

63. Fiedler, M.A., Wernke-Dollries, K. & Stark, J.M. Inhibition of TNF-alpha-induced NF-kappaB activation and IL-8 release in A549 cells with the proteasome inhibitor MG-132. Am J Respir Cell Mol Biol 19, 259-68 (1998).

64. Standiford, T.J. et al. Interleukin-8 gene expression by a pulmonary epithelial cell line. A model for cytokine networks in the lung. J Clin Invest 86, 1945-53 (1990).

65. Abraham, S.M. et al. Antiinflammatory effects of dexamethasone are partly dependent on induction of dual specificity phosphatase 1. J Exp Med 203, 1883-9 (2006).

66. Chi, H. et al. Dynamic regulation of pro- and anti-inflammatory cytokines by MAPK phosphatase 1 (MKP-1) in innate immune responses. Proc Natl Acad Sci U S A 103, 2274-9 (2006).

67. Hammer, M. et al. Control of dual-specificity phosphatase-1 expression in activated macrophages by IL-10. Eur J Immunol 35, 2991-3001 (2005).

68. Hammer, M. et al. Dual specificity phosphatase 1 (DUSP1) regulates a subset of LPS-induced genes and protects mice from lethal endotoxin shock. J Exp Med 203, 15-20 (2006).

69. Zhao, Q. et al. MAP kinase phosphatase 1 controls innate immune responses and suppresses endotoxic shock. J Exp Med 203, 131-40 (2006).

70. Sahu, B. et al. Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer. The EMBO journal 30, 3962-76 (2011).

71. Shi, Y. et al. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell 119, 941-53 (2004).

72. Gordon, M. et al. Genome-wide dynamics of SAPHIRE, an essential complex for gene activation and chromatin boundaries. Mol Cell Biol 27, 4058-69 (2007).

73. Choi, H.K. et al. Cohesion establishment factor, Eco1 represses transcription via association with histone demethylase, LSD1. Biochem Biophys Res Commun 394, 1063-8.

74. Lee, S.W. et al. ASXL1 represses retinoic acid receptor-mediated transcription through associating with HP1 and LSD1. J Biol Chem 285, 18-29.

75. Mulligan, P. et al. A SIRT1-LSD1 corepressor complex regulates Notch target gene expression and development. Mol Cell 42, 689-99.

76. van Riel, B. et al. A novel complex, RUNX1-MYEF2, represses hematopoietic genes in erythroid cells. Mol Cell Biol 32, 3814-22.

77. Yang, P. et al. RCOR2 is a subunit of the LSD1 complex that regulates ESC property and substitutes for SOX2 in reprogramming somatic cells to pluripotency. Stem Cells 29, 791-801.

78. Yatim, A. et al. NOTCH1 nuclear interactome reveals key regulators of its transcriptional activity and oncogenic function. Mol Cell 48, 445-58.

79. Lim, P.S. et al. Defining the chromatin signature of inducible genes in T cells. Genome biology 10, R107 (2009).

80. Wang, D. et al. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature 474, 390-4 (2011).

81. Zhou, B.O. & Zhou, J.Q. Recent transcription-induced histone H3 lysine 4 (H3K4) methylation inhibits gene reactivation. The Journal of biological chemistry 286, 34770-6 (2011).

82. Kim, S.M. et al. Regulation of mouse steroidogenesis by WHISTLE and JMJD1C through histone methylation balance. Nucleic acids research 38, 6389-403 (2010).

83. Wolf, S.S., Patchev, V.K. & Obendorf, M. A novel variant of the putative demethylase gene, s-JMJD1C, is a coactivator of the AR. Archives of biochemistry and biophysics 460, 56-66 (2007).

84. Ballas, N. et al. Regulation of neuronal traits by a novel transcriptional complex. Neuron 31, 353-65 (2001).

85. Huang, P., Chandra, V. & Rastinejad, F. Structural overview of the nuclear receptor superfamily: insights into physiology and therapeutics. Annual review of physiology 72, 247-72 (2010).

86. Goya, L., Maiyar, A.C., Ge, Y. & Firestone, G.L. Glucocorticoids induce a G1/G0 cell cycle arrest of Con8 rat mammary tumor cells that is synchronously reversed by steroid withdrawal or addition of transforming growth factor-alpha. Molecular endocrinology 7, 1121-32 (1993).

87. Sanchez, I., Goya, L., Vallerga, A.K. & Firestone, G.L. Glucocorticoids reversibly arrest rat hepatoma cell growth by inducing an early G1 block in cell cycle progression. Cell growth & differentiation : the molecular biology journal of the American Association for Cancer Research 4, 215-25 (1993).

88. Baghdassarian, N. et al. Glucocorticoids induce G1 as well as S-phase lengthening in normal human stimulated lymphocytes: differential effects on cell cycle regulatory proteins. Experimental cell research 240, 263-73 (1998).

89. Huang, Y., Johnson, K.R., Norris, J.S. & Fan, W. Nuclear factor-kappaB/IkappaB signaling pathway may contribute to the mediation of paclitaxel-induced apoptosis in solid tumor cells. Cancer research 60, 4426-32 (2000).

90. Pae, H.O., Yoo, J.C., Choi, B.M., Kim, T.Y. & Chung, H.T. Protective effects of glucocorticoids on taxol-induced cytotoxicity in human leukemia HL-60 cells. Immunopharmacology and immunotoxicology 21, 439-53 (1999).

91. Wolff, J.E., Denecke, J. & Jurgens, H. Dexamethasone induces partial resistance to cisplatinum in C6 glioma cells. Anticancer research 16, 805-9 (1996).

92. Uozumi, J., Koikawa, Y., Yasumasu, T., Tokuda, N. & Kumazawa, J. The protective effect of methylprednisolone against cisplatin-induced nephrotoxicity in patients with urothelial tumors. International journal of urology : official journal of the Japanese Urological Association 3, 343-7 (1996).

93. Herr, I. et al. Glucocorticoid cotreatment induces apoptosis resistance toward cancer therapy in carcinomas. Cancer research 63, 3112-20 (2003).

94. Shi, Y. et al. Coordinated histone modifications mediated by a CtBP co-repressor complex. Nature 422, 735-8 (2003).

95. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome biology 9, R137 (2008).

96. Shi, Y.J. et al. Regulation of LSD1 histone demethylase activity by its associated factors. Mol Cell 19, 857-64 (2005).

Chapter 3

H3K79 Histone Demethylation by JMJD2 family proteins.

Erin A. Clark[1], Paco Kang[1], and Yujiang Geno Shi[1]

1. Division of Endocrinology, Diabetes and Hypertension, Departments of Medicine and BCMP, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA

Author Contributions

The work presented in this chapter was conceptualized by Dr. Yujiang Shi and Erin Clark. The molecular and cellular experimentation was primarily conducted by Erin Clark. Paco Kang contributed significantly to the cloning of JMJD2C and generating the catalytic mutant as well as immunoflourescent photography.

Abstract

The field of histone demethylases has expanded rapidly following the discovery of the first histone lysine demehtylase, LSD1/KDM1A. New histone modifications are still being discovered along with the enzymes that add and remove these modifications. While the methyltransferases responsible for adding histone methylation have been characterized for all known sites of histone methylation, one site still remains without a characterized demethylase. H3K79 is an interesting site for modification given its location within the globular domain of the histone protein, unlike most modifications occurring on the unstructured tail. While DOT1 is known to methylate H3K79, no demethylase for this residue has been discovered to date. We present evidence here supporting never before characterized H3K79me3 demethylase activity by members of the JMJD2 family of proteins.

Background

HDMs are divided into two major families. One, the flavin-dependent LSD family consisting of KDM1A/LSD1 and KDM1B/LSD2, and two, the larger and more diverse Jumonji C (JmjC) family consisting of roughly 30 enzymes [1]. JmjC HDMs are Fe(II)-dependent hydroxylases that utilize the cosubstrates 2-oxoglutarate (2-OG) and molecular oxygen to catalyze demethylation. Several subfamilies of JmjC HDMs have been identified and characterized with distinct substrate specificities (Table 1.2) [2-9]. JmjC HDMs have been implicated in regulating diverse genomic processes, such as transcription, cell-cycle progression, heterochromatin maintenance, X chromosome inactivation, and development [3,5-7]. These functional roles vary among subfamilies and are believed to be a consequence of differing histone substrate specificities.

The JMJD2 subfamily is conserved from budding yeast to mammals [3,10-16]. Humans have four JMJD2 homologs (JMJD2A, JMJD2B, JMJD2C, and JMJD2D), which display dual selectivity in removing H3K9me2/3, a mark associated with heterochromatin and gene silencing, and H3K36me2/3, a modification demarcating the coding regions of actively transcribed genes [5,16]. JMJD2 HDMs have been implicated in transcriptional regulation, cell-cycle progression, nuclear hormone signaling, embryonic stem cell self-renewal, and development [5-7,17-22]. In addition, overexpression of several JMJD2 homologs has been linked to cancer [23-25].

While over 30 HDMs have been characterized in the less than 10 years since the discovery of the first LSD1, discovery of new HDMs is ongoing. The first histone arginine demethylase, JMJD6, was described as recently as 2007 [26], opening the field of HDMs beyond lysine residues for the

first time. The first H4K20me1 demethylase, PHF8, was only described in 2010 [27], leaving H4K20me2/3 without a known demethylase. Similarly one of only two methylation site known to occur on the histone globular domain, H3K79, has a characterized HMT, DOT1, but to date, no known cognate HDM.

DOT1 and H3K79 methylation have been implicated in regulating heterochromatin formation at telomeres [28] and active transcription. Genome-wide analysis finds that H3K79 methylation is enriched in the body of transcribed genes and the amount of methylation is correlated with expression level [29]. Similar studies in the yeast, fly, and human genomes strongly suggest that H3K79 methylation is a marker of active transcription [30-36]. The location of H3K79 methylation within the gene body along with findings that DOT1 is part of several complexes associated with elongating RNA polymerase II implicates this mark in regulating transcriptional elongation [37-39].

DOT1 and H3K79 methylation have also been shown to play an important role in DNA repair. The tandem tudor domain of the 53BP1 protein is able to bind methylated H3K79 and this function is required for its recruitment to DNA double-strand breaks [40,41]. Because H3K79 methylation levels do not change after DNA damage, it is thought that double-strand breaks change the chromatin structure to expose methylated H3K79, which is then recognized by 53BP1 [40]. DOT1-mediated H3K79 methylation has also been implicated in other forms of DNA repair such as nucleotide excision repair (NER), recombination repair (RR), or post-replication repair (PRR) and loss of DOT1 results in UV hypersensitivity [42,43]. Clearly the role of DOT1 and H3K79 methylation is complex and touches multiple important processes from heterochromatin formation at telomeres and transcription, to maintaining genome integrity. Each of these roles

potentially contributes to the biological functions in development and cardiac function attributed to DOT1 thus far [44,45]. They also implicate regulation of H3K79 methylation in important human diseases such as cancer (reviewed in [46]).

Although the establishment of H3K79 methylation is well characterized and a highly regulated processes, active demethylation of this mark is not well understood. However, several lines of evidence show dynamaic regulation of this mark therefore suggesting that H3K79 methylation is reversible. In yeast and human cells H3K79me2 levels fluctuate throughout the cell cycle [47,48], and in both mice and flies H3K79me2 is lost during early embryonic development [49,50]. Lastly, a study investigating the antagonistic effects of 2-hydroxyglutarate (2-HG) on the JmjC cofactor, 2-OG, showed a global increase in H3K79me2 suggesting that a dioxygenase such as a JmjC enzyme may catalyze the removal of H3K79me2 marks *in vivo* [51].

Here we show that overexpression of several JMJD2 family HDMs in cell culture results in a global decrease in H3K79me3 levels. Although more is required to fully establish JMJD2A, C, and D as bona fide direct H3K79me3 demethylases, these results support dynamic regulation of H3K79 methylation.

Results

JMJD2 family HDMs have been previously described as H3K36 and H3K9 demethylases. During the course of characterizing a putative HDM, JMJD2C was used as a positive control and found to have a previously uncharacterized activity against H3K79me3. Full-length, HA-tagged JMJD2C was overexpressed in the osteoblast cell line U2OS and

immunoflourescence staining against various histone marks were analyzed. A catalytically inactive mutant form of JMJD2C, M-JMJD2C, which disrupts two key amino acids in the iron binding domain (H189A and E191A) was used as a negative control. A global decrease in H3K79me3 in cells overexpressing wild-type JMJD2C but not the catalytic mutant (M-JMJD2C) was observed (Figure 3.1). Loss of H3K79me3 is specific to the trimethylated state as no decrease in H3K79me2 was observed (Figure 3.2). Demethylation of H3K36me3, a

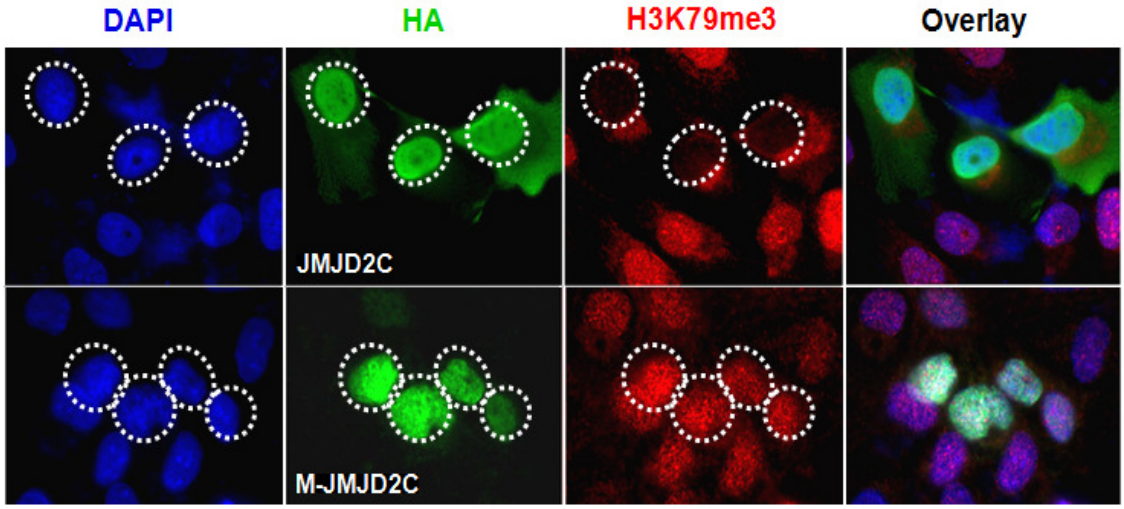


Figure 3.1 JMJD2C overexpression results in loss of H3K79me3 and is dependent on catalytic HDM activity.

HA-tagged, full-length JMJD2C or a catalytically inactive mutant, M-JMJD2C, was overexpressed in U2OS cells for 72 hours prior to immunoflourescent staining against H3K79me3. White dotted circles indicate transected nuclei. Global loss of H3K79me3 is observed in the wild-type JMJD2C transfected cells (top row), but not in the mutant transfected cells (bottom row).

known JMJD2 family substrate, was observed as a positive control (Figure 3.3 A), and as a negative control no demethylation of H3K4me3 was observed (Figure 3.3 B). Quantification of immunoflourescent intensity from histone marks showed that JMJD2C transfected cells

have significantly lower levels of H3K79me3 and H3K36me3 with a p-value ≤ 0.001 while JMJD2C-M transfected cells are not significantly different (Figure 3.4).

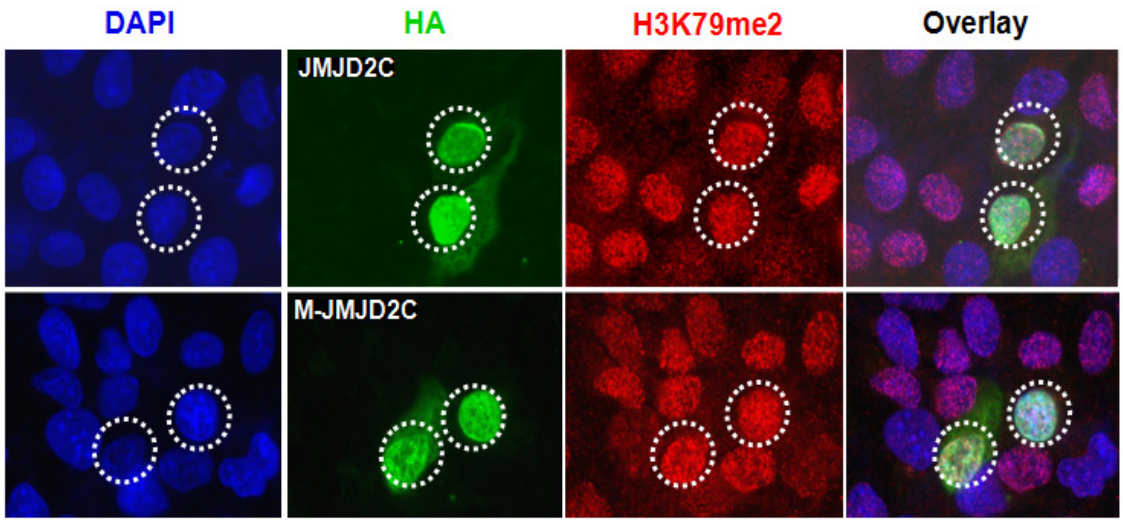


Figure 3.2 JMJD2C overexpression does not result in loss of H3K79me2.
HA-tagged, full-length JMJD2C or mutant, M-JMJD2C, was overexpressed in U2OS cells for 72 hours prior to immunoflourescent staining against H3K79me2. White dotted circles indicate transected nuclei. No detectible change in H3K79me2 levels is observed in either the wild-type or mutant transfected cells.

Given that the JMJD2 subfamily members share similar substrate specificities, two other members, JMJD2A and JMJD2D, were also tested. Overexpression of a truncated JMJD2A (1-310aa) containing the catalytic domain, and full-length JMJD2D also caused loss of H3K79me3 in U2OS cells as shown in Figure 3.5. This suggests H3K79me3 may be a substrate for the JMJD2 subfamily of HDMs. Initial experiments with the lung adenocarcenoma cell line, A549, suggests this result is not limited to U2OS cells (data not shown).

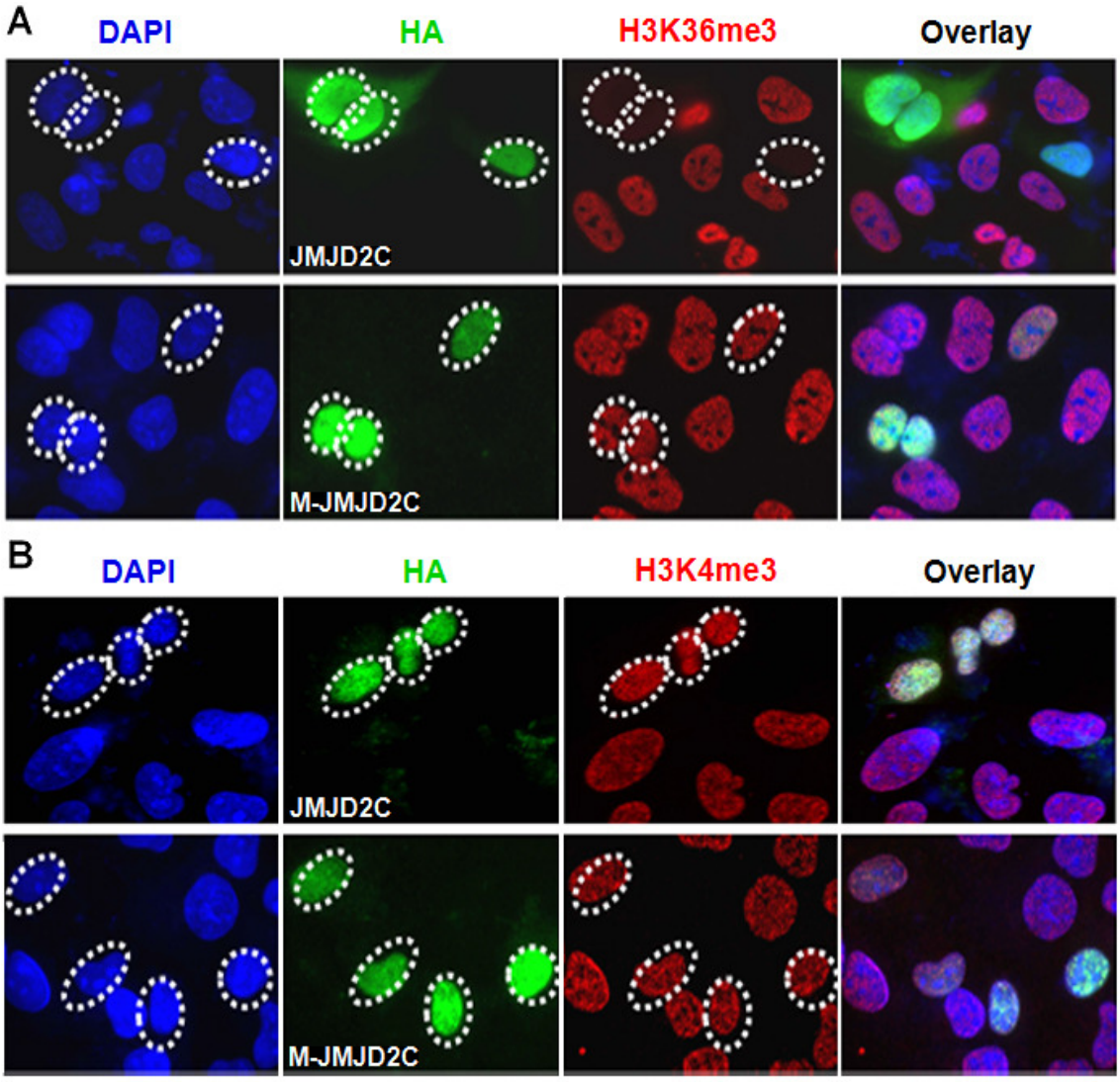


Figure 3.3 In addition to of loss of H3K79me3, H3K36me3 is also lost while H3K4me3 is not. HA-tagged, full-length JMJD2C or mutant, M-JMJD2C, was overexpressed in U2OS cells for 72 hours prior to immunoflourescent staining against H3K36me3 (A), or H3K4me3 (B). White dotted circles indicate transected nuclei. Global loss of H3K36me3 is observed in the wild-type JMJD2C transfected cells (top row of A), but not in the mutant transfected cells (bottom row of A). No detectible change in H3K4me3 levels is observed in either the wild-type or mutant transfected cells.

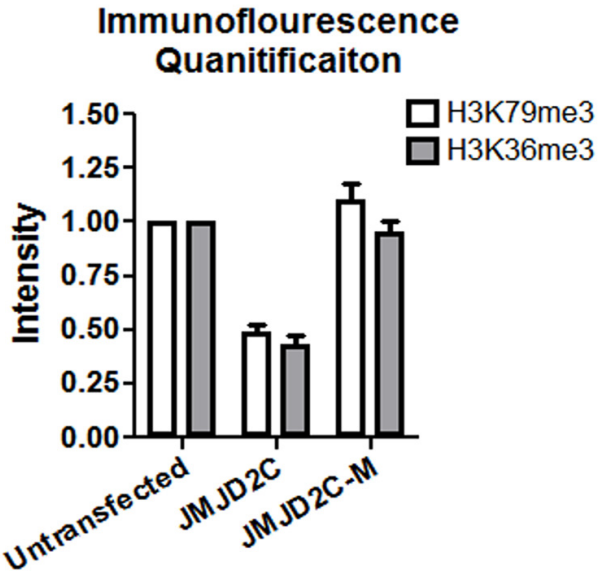


Figure 3.4 Quantification of overexpressed JMJD2C activity on H3K79me3 and H3K36me3.

Immunoflourescence intensity of histone modifications were quantified in cells expressing either wild-type or mutant JMJD2C and compared to neighboring untransfected cells. JMJD2C transfected cells have significantly lower levels of H3K79me3 and H3K36me3 with a p-value ≤ 0.001 while JMJD2C-M transfected cells are not significantly different. N = 10-20 cells from two separate experiments.

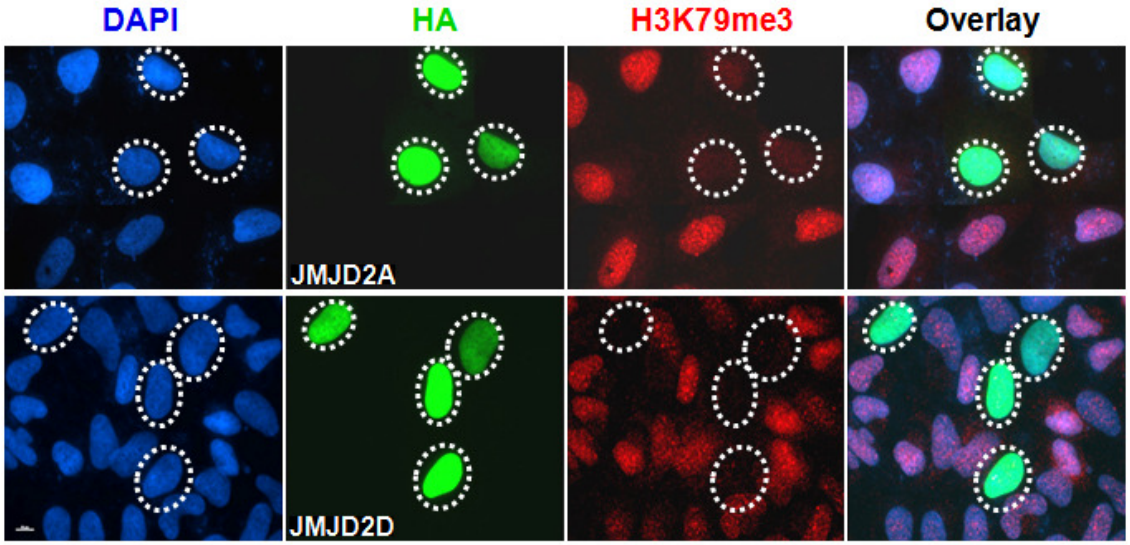


Figure 3.5 Overexpression of other JMJD2 family members, JMJD2A and JMJD2D, also result in loss of H3K79me3.

HA-tagged, truncated JMJD2A (1-310aa) or HA-tagged, full-length JMJD2D was overexpressed in U2OS cells for 72 hours prior to immunoflourescent staining against H3K79me3. White dotted circles indicate transected nuclei. Global loss of H3K79me3 is observed in cells transfected with either construct.

Discussion

Since the first histone demethylase was discovered in 2004, a number of enzymes have been characterized as histone demethylases acting on specific histone residuals. However, no enzyme has of yet been identified with H3K79 demethylase activity. The evidence presented here suggests that the JMJD2 subfamily of JmjC HDMs may act as H3K79me3 demethylases. The observation that mutating the catalytic domain of JMJD2C abolishes loss of H3K79me3 upon overexpression strongly suggests H3K79me3 levels are regulated by JMJD2C enzymatic activity. While we cannot yet rule out the possibility of indirect regulation via demethylation of other JMJD2 histone substrates, no relationship between

H3K36 or H3K9 methylation, two well-established JMJD2 substrates, has been described that could account for the observed H3K79me3 loss. Further *in vitro* and biochemical characterization with purified JMJD2 proteins will be necessary to convincingly established H3K79me3 as a bona fide substrate.

One possible explanation for the lag in discovering a H3K79 demethylase may be due to unique features related this mark occurring on the globular region of the histone protein rather than the tail. Previous studies have shown no activity of JMJD2 HDMs on methylated H3K79 peptides. However, characterization of DOTL enzymatic activity shows preferential methylation of H3K79 in the context of nucleosomes rather than core histones or recombinant H3 peptide [47], suggesting recognition of other features of the nucleosome beyond K79 and adjacent residues. This may very well be true for the reverse reaction recognizing and removing H3K79 methylation, making some of the standard techniques used to characterize novel HDMs insensitive to H3K79 demethylase activity.

Interestingly, the substrate specificity of several JMJD2 family members, JMJD2A-D, was recently expanded to include the newly discovered H3K56me3 mark [52]. Similar to H3K79, H3K56 is located on the globular core of the histone protein further supporting the possibility that JMJD2 proteins are capable of recognizing and acting on core residues.

H3K79 methylation has been implicated in various cellular processes such as DNA damage repair, cell cycle regulation, transcription and regulation of telomeric heterochromatin

formation [40,53,54]. It would be interesting to test whether JMJD2A/C/D have roles in these processes via their H3K79me3 HDM activity.

Experimental Procedures

U2OS cells were cultured in DMEM supplemented with 20%. For immunoflourescence, cells were fixed with 4% paraformaldehyde and permeabilized with 0.5% triton-X100. After blocking with BSA cells were incubated with the indicated antibody, except H3K79me3, for 2 hours at room temperature. The H3K79me3 antibody (abcam ab2621) was significantly more difficult to stain and therefore incubated for 6-8 hours at room temperature.

M-JMJD2C was generated by site directed mutagenesis of the iron binding site amino acids H189A and E191A.

Aknowledgments

We thank Johnathan Whetstine for kindly sharing JMJD2A and JMJD2D constructs with us.

References

1. Kooistra, S.M. & Helin, K. Molecular mechanisms and potential functions of histone demethylases. Nat Rev Mol Cell Biol 13, 297-311.

2. Allis, C.D. et al. New nomenclature for chromatin-modifying enzymes. Cell 131, 633-6 (2007).

3. Cloos, P.A., Christensen, J., Agger, K. & Helin, K. Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. Genes & development 22, 1115-40 (2008).

4. Krishnan, S., Horowitz, S. & Trievel, R.C. Structure and Function of Histone H3 Lysine 9 Methyltransferases and Demethylases. Chembiochem : a European journal of chemical biology (2011).

5. Mosammaparast, N. & Shi, Y. Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases. Annual review of biochemistry 79, 155-79 (2010).

6. Nottke, A., Colaiacovo, M.P. & Shi, Y. Developmental roles of the histone lysine demethylases. Development 136, 879-89 (2009).

7. Shi, Y. Histone lysine demethylases: emerging roles in development, physiology and disease. Nature reviews. Genetics 8, 829-33 (2007).

8. Shi, Y. & Whetstine, J.R. Dynamic regulation of histone lysine methylation by demethylases. Molecular cell 25, 1-14 (2007).

9. Upadhyay, A.K., Horton, J.R., Zhang, X. & Cheng, X. Coordinated methyl-lysine erasure: structural and functional linkage of a Jumonji demethylase domain and a reader domain. Current opinion in structural biology 21, 750-60 (2011).

10. Fodor, B.D. et al. Jmjd2b antagonizes H3K9 trimethylation at pericentric heterochromatin in mammalian cells. Genes & development 20, 1557-62 (2006).

11. Katoh, M. Identification and characterization of JMJD2 family genes in silico. International journal of oncology 24, 1623-8 (2004).

12. Kim, T. & Buratowski, S. Two Saccharomyces cerevisiae JmjC domain proteins demethylate histone H3 Lys36 in transcribed regions to promote elongation. The Journal of biological chemistry 282, 20827-35 (2007).

13. Klose, R.J. et al. The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. Nature 442, 312-6 (2006).

14. Klose, R.J. et al. Demethylation of histone H3K36 and H3K9 by Rph1: a vestige of an H3K9 methylation system in Saccharomyces cerevisiae? Molecular and cellular biology 27, 3951-61 (2007).

15. Tu, S. et al. Identification of histone demethylases in Saccharomyces cerevisiae. The Journal of biological chemistry 282, 14262-71 (2007).

16. Whetstine, J.R. et al. Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. Cell 125, 467-81 (2006).

17. Black, J.C. et al. Conserved antagonism between JMJD2A/KDM4A and HP1gamma during cell cycle progression. Molecular cell 40, 736-48 (2010).

18. Iwamori, N., Zhao, M., Meistrich, M.L. & Matzuk, M.M. The testis-enriched histone demethylase, KDM4D, regulates methylation of histone H3 lysine 9 during spermatogenesis in the mouse but is dispensable for fertility. Biology of reproduction 84, 1225-34 (2011).

19. Lin, C.H. et al. Heterochromatin protein 1a stimulates histone H3 lysine 36 demethylation by the Drosophila KDM4A demethylase. Molecular cell 32, 696-706 (2008).

20. Loh, Y.H., Zhang, W., Chen, X., George, J. & Ng, H.H. Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. Genes & development 21, 2545-57 (2007).

21. Lorbeck, M.T. et al. The histone demethylase Dmel\Kdm4A controls genes required for life span and male-specific sex determination in Drosophila. Gene 450, 8-17 (2010).

22. Strobl-Mazzulla, P.H., Sauka-Spengler, T. & Bronner-Fraser, M. Histone demethylase JmjD2A regulates neural crest specification. Developmental cell 19, 460-8 (2010).

23. Shin, S. & Janknecht, R. Activation of androgen receptor by histone demethylases JMJD2A and JMJD2D. Biochemical and biophysical research communications 359, 742-6 (2007).

24. Wissmann, M. et al. Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression. Nature cell biology 9, 347-53 (2007).

25. Kim, T.D., Oh, S., Shin, S. & Janknecht, R. Regulation of tumor suppressor p53 and HCT116 cell physiology by histone demethylase JMJD2D/KDM4D. PloS one 7, e34618 (2012).

26. Chang, B., Chen, Y., Zhao, Y. & Bruick, R.K. JMJD6 is a histone arginine demethylase. Science 318, 444-7 (2007).

27. Qi, H.H. et al. Histone H4K20/H3K9 demethylase PHF8 regulates zebrafish brain and craniofacial development. Nature 466, 503-7.

28. Norris, A. & Boeke, J.D. Silent information regulator 3: the Goldilocks of the silencing complex. Genes & development 24, 115-22 (2010).

29. Steger, D.J. et al. DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. Molecular and cellular biology 28, 2825-39 (2008).

30. Pokholok, D.K. et al. Genome-wide map of nucleosome acetylation and methylation in yeast. Cell 122, 517-27 (2005).

31. Ng, H.H., Ciccone, D.N., Morshead, K.B., Oettinger, M.A. & Struhl, K. Lysine-79 of histone H3 is hypomethylated at silenced loci in yeast and mammalian cells: a potential mechanism for position-effect variegation. Proceedings of the National Academy of Sciences of the United States of America 100, 1820-5 (2003).

32. Schubeler, D. et al. The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes & development 18, 1263-71 (2004).

33. Okada, Y. et al. hDOT1L links histone methylation to leukemogenesis. Cell 121, 167-78 (2005).

34. Okada, Y. et al. Leukaemic transformation by CALM-AF10 involves upregulation of Hoxa5 by hDOT1L. Nature cell biology 8, 1017-24 (2006).

35. Vakoc, C.R., Sachdeva, M.M., Wang, H. & Blobel, G.A. Profile of histone lysine methylation across transcribed mammalian chromatin. Molecular and cellular biology 26, 9185-95 (2006).

36. Wang, Z. et al. Combinatorial patterns of histone acetylations and methylations in the human genome. Nature genetics 40, 897-903 (2008).

37. Krogan, N.J. et al. The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. Molecular cell 11, 721-9 (2003).

38. Mueller, D. et al. Misguided transcriptional elongation causes mixed lineage leukemia. PLoS biology 7, e1000249 (2009).

39. Bitoun, E., Oliver, P.L. & Davies, K.E. The mixed-lineage leukemia fusion partner AF4 stimulates RNA polymerase II transcriptional elongation and mediates coordinated chromatin remodeling. Human molecular genetics 16, 92-106 (2007).

40. Huyen, Y. et al. Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. Nature 432, 406-11 (2004).

41. Wysocki, R. et al. Role of Dot1-dependent histone H3 methylation in G1 and S phase DNA damage checkpoint functions of Rad9. Molecular and cellular biology 25, 8430-43 (2005).

42. Bostelman, L.J., Keller, A.M., Albrecht, A.M., Arat, A. & Thompson, J.S. Methylation of histone H3 lysine-79 by Dot1p plays multiple roles in the response to UV damage in Saccharomyces cerevisiae. DNA repair 6, 383-95 (2007).

43. Lazzaro, F. et al. Histone methyltransferase Dot1 and Rad9 inhibit single-stranded DNA accumulation at DSBs and uncapped telomeres. The EMBO journal 27, 1502-12 (2008).

44. Nguyen, A.T. et al. DOT1L regulates dystrophin expression and is critical for cardiac function. Genes Dev 25, 263-74.

45. Jones, B. et al. The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure. PLoS Genet 4, e1000190 (2008).

46. Nguyen, A.T. & Zhang, Y. The diverse functions of Dot1 and H3K79 methylation. Genes & development 25, 1345-58 (2011).

47. Feng, Q. et al. Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. Current biology : CB 12, 1052-8 (2002).

48. Schulze, J.M. et al. Linking cell cycle to histone modifications: SBF and H2B monoubiquitination machinery and cell-cycle regulation of H3K79 dimethylation. Molecular cell 35, 626-41 (2009).

49. Shanower, G.A. et al. Characterization of the grappa gene, the Drosophila histone H3 lysine 79 methyltransferase. Genetics 169, 173-84 (2005).

50. Ooga, M. et al. Changes in H3K79 methylation during preimplantation development in mice. Biology of reproduction 78, 413-24 (2008).

51. Xu, W. et al. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. Cancer cell 19, 17-30 (2011).

52. Jack, A.P. et al. H3K56me3 is a novel, conserved heterochromatic mark that largely but not completely overlaps with H3K9me3 in both regulation and localization. PloS one 8, e51765 (2013).

53. Schulze, J.M. et al. Linking cell cycle to histone modifications: SBF and H2B monoubiquitination machinery and cell-cycle regulation of H3K79 dimethylation. Mol Cell 35, 626-41 (2009).

54. Steger, D.J. et al. DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. Mol Cell Biol 28, 2825-39 (2008).

Appendix A

Molecular Cell
Article

LSD2/KDM1B and Its Cofactor NPAC/GLYR1 Endow a Structural and Molecular Model for Regulation of H3K4 Demethylation

Rui Fang,[1,8] Fei Chen,[2,4,8] Zhenghong Dong,[2,4] Di Hu,[1,2] Andrew J. Barbera,[1] Erin A. Clark,[1] Jian Fang,[2,4] Ying Yang,[2,4] Pinchao Mei,[1] Michael Rutenberg,[1] Ze Li,[2,4] Ying Zhang,[2,3] Youwei Xu,[2,4] Huirong Yang,[2,4] Ping Wang,[2,4] Matthew D. Simon,[7] Qiongjie Zhou,[1,5] Jing Li,[1] Mark P. Marynick,[1] Xiaotian Li,[5] Haojie Lu,[2,3] Ursula B. Kaiser,[1] Robert E. Kingston,[7] Yanhui Xu,[2,4,*] and Yujiang Geno Shi[1,2,6,*]

[1]Division of Endocrinology, Diabetes, and Hypertension, Department of Medicine and Department of Biological Chemistry & Molecular Pharmacology, Brigham and Women's Hospital and Harvard Medical School, 221 Longwood Avenue, Boston, MA 02115, USA
[2]Institutes of Biomedical Sciences
[3]Department of Chemistry
[4]State Key Laboratory of Genetic Engineering
[5]Obstetrics and Gynecology Hospital
[6]Children's Hospital
Fudan University, 130 Dong An Road, Shanghai 200032, China
[7]Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA
[8]These authors contributed equally to this work
*Correspondence: xuyh@fudan.edu.cn (Y.X.), yujiang_shi@hms.harvard.edu (Y.G.S.)
http://dx.doi.org/10.1016/j.molcel.2012.11.019

SUMMARY

Dynamic regulation of histone methylation represents a fundamental epigenetic mechanism underlying eukaryotic gene regulation, yet little is known about how the catalytic activities of histone demethylases are regulated. Here, we identify and characterize NPAC/GLYR1 as an LSD2/KDM1b-specific cofactor that stimulates H3K4me1 and H3K4me2 demethylation. We determine the crystal structures of LSD2 alone and LSD2 in complex with the NPAC linker region in the absence or presence of histone H3 peptide, at resolutions of 2.9, 2.0, and 2.25 Å, respectively. These crystal structures and further biochemical characterization define a dodecapeptide of NPAC (residues 214–225) as the minimal functional unit for its cofactor activity and provide structural determinants and a molecular mechanism underlying the intrinsic cofactor activity of NPAC in stimulating LSD2-catalyzed H3K4 demethylation. Thus, these findings establish a model for how a cofactor directly regulates histone demethylation and will have a significant impact on our understanding of catalytic-activity-based epigenetic regulation.

INTRODUCTION

Since the discovery of the first histone-lysine-specific demethylase, LSD1/KDM1a, histone-lysine demethylation has emerged as an epigenetic paradigm (Shi et al., 2004). Thus far, over 20 histone-lysine demethylases (KDMs) have been characterized, belonging to either the flavin adenine dinucleotide (FAD)-dependent LSD family or the Fe^{2+} and α-ketoglutarate-dependent Jumonji C-terminal domain family (Allis et al., 2007; Bernstein et al., 2007; Chen et al., 2006b; Tsukada et al., 2006; Rice and Allis, 2001; Ruthenburg et al., 2007). Genetic, biochemical, and functional studies further indicate that these KDMs play crucial roles in a wide range of biological processes, including gene expression, cell growth, differentiation, development, and disease pathogenesis (Bhaumik et al., 2007; Egger et al., 2004; Esteller, 2008; Nottke et al., 2009; Shi, 2007).

A key question remains in the mechanistic understanding of precisely how the enzymatic activities of KDMs are regulated (Chen et al., 2006b; Horton et al., 2010; Lan et al., 2008; Wilson, 2007). It has been observed that many KDMs, although being active on synthetic peptides or core histone substrates, exhibit very weak or no detectable activity on nucleosomal substrates in vitro. When transfected into cells, however, robust activity on chromatin can be detected, suggesting the existence of additional cofactors required for full activity (Shi et al., 2004; Tahiliani et al., 2007). We, and others, have identified CoREST as a cofactor required for LSD1 action on nucleosomal substrates (Lee et al., 2005; Shi et al., 2005), representing the first breakthrough toward an understanding of how KDM activity is regulated. However, the molecular details underlying the cofactor-enhanced demethylase activity of LSD1 remain elusive (Forneris et al., 2007; Yang et al., 2006). Moreover, the cofactor activity of CoREST is highly specific, facilitating only demethylation of nucleosomal substrates by LSD1, but not any other KDMs. Further investigation is required for determining whether cofactor modulation is a general mechanism for the regulation of KDM functions. In particular, there are two areas to be addressed: first, whether different cofactors exist for other histone demethylases; and second, the molecular mechanism(s) employed by such cofactors to facilitate histone demethylase activity.

LSD2/KDM1b/AOF1 is the only mammalian ortholog of LSD1 and possesses similar histone H3K4 demethylase activity (Ciccone et al., 2009; Fang et al., 2010; Yang et al., 2010). However, LSD2 is a component of a different cellular complex and has functions distinct from those of LSD1 (Ciccone et al., 2009; Fang et al., 2010; van Essen et al., 2010). Genetic studies indicate that LSD2 is required for the homeostasis of global H3K4 methylation in mouse oocytes and regulates maternal gene imprinting (Ciccone et al., 2009). In somatic tissue, LSD2 seems to play an important role in active gene transcription. LSD2 is reported to be a potential H3K9 demethylase and is required for controlling NF-κB-induced gene activation by demethylating H3K9 at promoters (van Essen et al., 2010). On the other hand, we show that LSD2 is an active H3K4 demethylase that specifically associates with the coding region of target genes. Removal of endogenous LSD2 promotes an increase in H3K4me2 levels and a concurrent decrease in H3K9me2 levels specifically at coding regions, but not at the corresponding promoters, and results in downregulation of gene transcription (Fang et al., 2010). These genetic and functional studies suggest that LSD2 is an important epigenetic regulator involved in diverse biological processes. How LSD2 activity is targeted to various functional sites and whether its activity is regulated by specific cofactors remain unknown.

Here, we report a cofactor of LSD2, NPAC/GLYR1, which positively regulates H3K4me2- and H3K4me1-specific histone demethylase activity of LSD2. NPAC, a putative H3K36me3 reader (Vermeulen et al., 2010), is a previously uncharacterized integral component of the LSD2 histone demethylase complex (Fang et al., 2010). We show that NPAC directly interacts with LSD2 and positively regulates its H3K4 demethylation activity both in vitro and in vivo. To understand the precise molecular mechanism of NPAC in regulating LSD2 enzymatic activity, we determined the crystal structures of LSD2, LSD2 in complex with NPAC, and the ternary complex of LSD2-NPAC-H3 peptide. These structural studies, together with molecular and biochemical characterization, illustrate a molecular model of cofactor-mediated regulation of the catalytic activity of a histone demethylase.

RESULTS

NPAC Is a Cofactor of LSD2 Positively Regulating Its H3K4 Histone Demethylase Activity

NPAC/GLYR1 (UniProtKB/Swiss-Prot ID: Q49A26) contains multiple functional domains, including a PWWP (Pro-Trp-Trp-Pro) domain, an AT-hook motif, and a dehydrogenase domain (Figure 1A). The presence of a dehydrogenase domain within NPAC was particularly intriguing because CtBP, a well-known corepressor and component of the LSD1 complex, also possesses a dehydrogenase domain (Chinnadurai, 2007; Shi et al., 2003). The potential analogy of NPAC/LSD2 to CtBP/LSD1 prompted us to focus on understanding the activity of NPAC in relation to LSD2 function. However, unlike CtBP, attempts to identify the intrinsic enzymatic activity of NPAC as a potential dehydrogenase using either recombinant NPAC protein purified from *E. coli* or the NPAC complex purified from HeLa cells via tandem affinity purification (TAP) were unsuccessful (data not shown).

Although no intrinsic histone demethylase activity was observed for recombinant NPAC, we did observe that the NPAC complex purified from HeLa cells via TAP has robust H3K4me2 demethylase activity toward nucleosomes (Figure 1B, lane 1). The H3K4 demethylase activity of the NPAC complex is probably attributable to endogenous LSD2, given that LSD2 was the only histone demethylase detected in the complex by mass spectrometry (MS; data not shown). Paradoxically, although the H3K4 demethylase activity was higher, the relative amount of LSD2 in the NPAC complex was significantly lower than that of the LSD2 complex, in which a small amount of endogenous NPAC was copurified with LSD2 (Figure 1Bg). The addition of purified recombinant NPAC to the LSD2 complex significantly increased nucleosomal demethylation (Figure 1C). These observations suggest that NPAC may positively regulate LSD2 histone demethylase activity on nucleosomes, similar to the cofactor activity of CoREST for LSD1 (Lee et al., 2005; Shi et al., 2005).

To directly validate its cofactor activity, we investigated whether NPAC alone is sufficient to enhance the activity of recombinant LSD2 in vitro. We have previously reported that, unlike LSD1, recombinant LSD2 can demethylate nucleosomal H3K4me2 at a high dosage (Fang et al., 2010). Therefore, to measure the stimulatory effect of NPAC on nucleosomal substrates, we titrated down the amount of recombinant hexa-His-tagged LSD2 (His-LSD2) until no obvious histone demethylase activity was detected (Figure S1A, lane 5, available online). Using this threshold dosage of His-LSD2, we performed nucleosome demethylation assays with increasing amounts of recombinant His-NPAC, which stimulated LSD2 activity in a dose-dependent manner (Figure 1D). In contrast, NPAC has no stimulatory effect on nucleosome demethylation by LSD1 (Figure S1B), and NPAC itself possesses no demethylase activity (Figure 1D, lane 9).

To further validate NPAC cofactor function in vivo, we coexpressed NPAC and LSD2, and we observed a much more pronounced depletion of di- and monomethylated H3K4 in cells than when expressing LSD2 alone (Figures 1E and S1C). Coexpression of NPAC does not change the substrate specificity of LSD2, because H3K9me2 and other histone marks examined showed no detectable changes (Figure S1D and data not shown). Taken together, both in vitro and in vivo data suggest that NPAC is a specific cofactor for LSD2, positively regulating its H3K4-specific histone demethylase activity.

The Linker Region of NPAC Is Sufficient for Cofactor Activity and LSD2 Interaction

To identify the functional domain responsible for NPAC cofactor activity, we generated NPAC deletion mutants (Figure 2A) and examined their abilities to facilitate LSD2 activity on nucleosome substrates. Neither the PWWP domain (NP.d1, residues 1–150) nor the dehydrogenase domain (NP.d5, residues 262–553) stimulated LSD2 (Figure 2B, lanes 5 and 9 compared to lanes 2 and 11). In contrast, truncation proteins containing the linker region, NP.d2 (residues 1–252), NP.d3 (residues 152–252), and NP.d4 (residues 152–553), significantly enhanced LSD2 demethylase activity (lanes 6–8), whereas deletion of the linker region (residues 152–252) abolished cofactor activity (NP.d6, lane 10). Furthermore, the linker region of NPAC exhibits strong cofactor

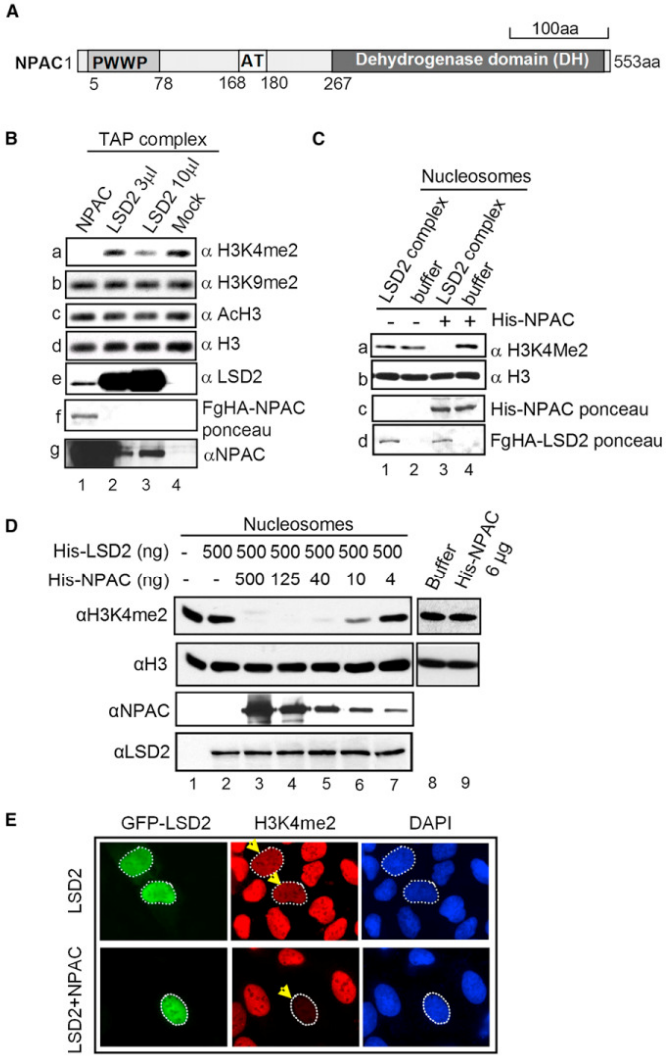


Figure 1. NPAC Is a Cofactor of LSD2 Positively Modulating Its H3K4 Demethylase Activity

(A) Schematic of NPAC domain structure. AT, AT-hook motif; PWWP, Pro-Trp-Trp-Pro domain.

(B) LSD2 in the NPAC complex can efficiently demethylate nucleosomal H3K4me2. Tandem-affinity-purified NPAC and LSD2 complexes were incubated with nucleosomes purified from HeLa cells and analyzed by immunoblot using the indicated antibodies.

(C) The addition of recombinant NPAC can improve H3K4 demethylase activity of the LSD2 complex.

(D) Recombinant NPAC stimulates LSD2 nucleosomal H3K4 demethylation in a dose-dependent manner. The amount of LSD2-enzyme and NPAC-cofactor proteins used in each reaction is indicated. Demethylation was assessed by immunoblot using the indicated antibodies. Significantly larger amounts of His-LSD2 are required for the efficient demethylation of nucleosomes (see Figure S1A).

(E) NPAC stimulates H3K4 demethylation mediated by LSD2 in cells. Immunofluorescence staining of U2OS transiently transfected with LSD2 alone or in combination with NPAC is shown. Green, GFP-LSD2; red, H3K4me2; blue, DAPI counterstain of DNA. Representatives with similar levels of GFP-LSD2 are shown. See also Figure S1.

activity for LSD2 when a synthetic H3K4me2 peptide (residues 1–21) is used as the substrate (Figure 2C).

Examining the same NPAC mutants using glutathione S-transferase (GST)-pull-down assays, we further showed that truncation proteins possessing strong cofactor activity (NP.d2, NP.d3, and NP.d4) are also capable of directly interacting with LSD2 (Figure 2D). Neither the PWWP domain (NP.d1) nor the dehydrogenase domain (NP.d5) interacted with LSD2 (lanes 4 and 8, respectively), and removal of either domain (NP.d4 or NP.d2) did not adversely affect LSD2 binding (lanes 7 and 5, compared to lane 3). Removing the linker region (NP.d6) abolished NPAC-LSD2 interaction (lane 9).

Taking these results together, we conclude that the linker region of NPAC is responsible for LSD2 interaction as well as for cofactor activity. It has been proposed that CoREST stimulates LSD1 histone demethylase activity on nucleosomes by assisting in the docking of LSD1-CoREST to nucleosomal substrates via its DNA-binding SANT2 domain (Yang et al., 2006). In contrast, the linker region of NPAC facilitates LSD2 enzymatic activity, regardless of whether the substrate is chromatin, nucleosome, or a modified short histone peptide, and the nucleosome-binding PWWP domain of NPAC is dispensable for its cofactor activity. These observations suggest a mechanism for the direct action of a cofactor on histone demethylation, not depending on the nucleosomal context, but primarily involving the histone tail.

The Crystal Structures of LSD2, LSD2-NPAC, and the Ternary Complex of LSD2-NPAC-H3 Peptide

To understand the structural basis for LSD2 function and its regulation by NPAC, we solved the LSD2 crystal structure using

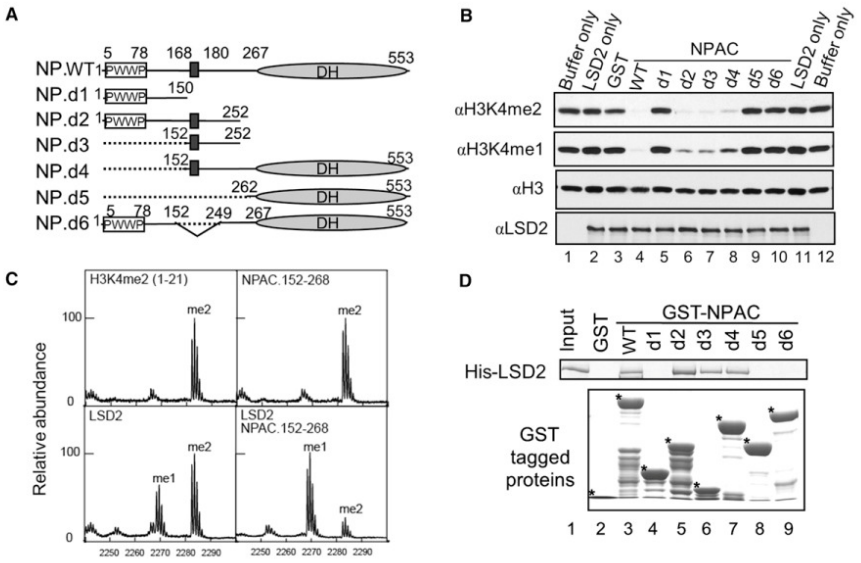


Figure 2. The Linker Region of NPAC Is Sufficient for Its Cofactor Activity and LSD2 Interaction

(A) Schematic representation of the wild-type and deletion mutants of NPAC. DH, dehydrogenase domain; black box, AT-hook motif.

(B) Nucleosome demethylation assays examining cofactor activities of NPAC mutants. Equal amounts of LSD2 were used in nucleosome demethylation reactions 2–11, in combination with various GST-tagged NPAC truncation proteins indicated above. GST was included as a negative control.

(C) The linker region of NPAC can stimulate LSD2 histone demethylase activity toward short H3K4me2 peptides. Molecular masses corresponding to mono- and dimethylated H3K4 peptides (residues 1–21) are denoted as me1 and me2, respectively.

(D) The linker region of NPAC is sufficient for LSD2 binding. Purified GST and GST-tagged wild-type and mutant NPAC proteins were used for GST pull-down of purified His-LSD2. Pull-down complexes were separated by SDS-PAGE and visualized by Coomassie blue staining. Asterisk, GST fusion protein.

a truncated protein (residues 51–822) purified to homogeneity, as well as the cocrystal structures of LSD2 in complex with the NPAC linker region (amino acids 152–268) in the presence and absence of H3K4M peptides (Histone H3 residues 1–21, replacing K4 with a methionine to mimic the H3K4me2 substrate of LSD2) (Forneris et al., 2007). Resolutions of the refined models were 2.9 Å, 2.25 Å, and 2.0 Å, respectively (Table 1, Table S1, and Figure S2).

The structure of LSD2 adopts a compact rod shape and comprises four recognizable domains: a zinc-finger domain with two zinc atoms coordinated by a $Cys_4His_2Cys_2$ motif (ZF, lime), a CW-type zinc-finger domain with one zinc atom coordinated by four cysteines (Zf-CW, purple), a SWIRM domain (red), and an amine oxidase domain (AO, green) (Figures 3A and 3B). Compared to the structure of LSD2 alone, the cocrystal structures show that NPAC binding does not significantly alter the overall structure of LSD2 or induce conformational changes in the catalytic domain. Similarly, the inclusion of histone H3K4M peptide does not lead to significant conformational changes in the LSD2-NPAC complex (Figures 3C, 3D, and S3A and Table S2). This suggests that, unlike regulators of the SET1-family histone methyltransferases (Dou et al., 2006; Southall et al., 2009), NPAC does not employ an allosteric mechanism to facilitate LSD2 enzymatic activity or switch LSD2 substrate specificity (Figures S3B and S3C).

The LSD2 Structure Reveals Features Common with and Distinct from LSD1

LSD2 and LSD1 share significant similarities in the AO catalytic domain (Chen et al., 2006a; Forneris et al., 2007; Stavropoulos et al., 2006; Yang et al., 2006). The overall folding and positions of the catalytic residue K661 and residues coordinating the FAD coenzyme are well conserved, creating indistinguishable catalytic cavities (Figures S4A–S4C), consistent with their similar substrate specificities (Ciccone et al., 2009; Fang et al., 2010; Shi et al., 2004; Yang et al., 2010). However, the crystal structure reveals several distinctive structural features of LSD2, which may significantly influence its intrinsic histone demethylase activity and explain differing regulatory mechanisms.

The most striking structural difference between LSD2 and LSD1 is the "tower domain," which is present in the LSD1 AO domain but absent in LSD2 (Figure 4A). The tower domain of LSD1 is the binding site for CoREST. The lack of a tower domain in LSD2 inherently necessitates an alternate mechanism for LSD2-cofactor interaction. Another distinction is that the two zinc-finger domains present in LSD2 are absent in LSD1 (Figures

Table 1. Crystallographic Data and Structure Refinement Statistics

Data Collection			
Crystal (PDB ID code)	LSD2 (4GU1)	LSD2-NPAC (4GUT)	LSD2-NPAC-H3K4M (4GUS)[c]
Wavelength (Å)	0.97947	0.97916	1.00001
Resolution (Å)	50.00–2.90 (3.00–2.90)	50.00–2.00 (2.07–2.00)	50.00–2.25 (2.33–2.25)
Space group	$P2_12_12_1$	$P2_1$	$P3_221$
Cell parameters (Å, °)	*a* = 89.2 *b* = 89.2 *c* = 342.5	*a* = 62.0 *b* = 89.8 *c* = 86.7 β = 105.0°	*a* = 101.1 *b* = 101.1 *c* = 177.4 γ = 120.0°
Completeness (%)	99.5 (98.2)	95.2 (75.9)	99.9 (100.0)
R_{merge} (%)	12.4 (67.0)	7.6 (34.6)	8.1 (65.3)
I/σ (I)	17.4 (2.5)	23.5 (3.6)	24.6 (3.6)
Redundancy	13.5 (8.6)	6.8 (5.0)	10.6 (10.6)
No. of all reflections	785,997 (48,478)	402,920 (23,610)	548,052 (54,314)
No. of unique reflections	58,222 (5,637)	59,253 (4,722)	51,703 (5,124)
Refinement Statistics			
Resolution (Å)[a]	50.00–2.90 (3.00–2.90)	50.00–2.00 (2.07–2.00)	50.00–2.25 (2.33–2.25)
R_{work}/R_{free} (%)[b]	18.50/22.79	19.60/20.72	20.71/23.70
Deviation from Identity			
Bonds (Å)	0.011	0.011	0.008
Angles (°)	1.342	1.147	1.039
Average B factor (Å2)	83.438	39.159	44.413
Ramachandran Plot Statistics			
Most favored regions (%)	85.8	91.3	88.2
Allowed regions (%)	13.8	8.4	11.5
Generously allowed regions (%)	0.1	0	0
Disallowed regions (%)[d]	0.3	0.3	0.3

PDB, Protein Data Bank.
[a]The values for the data in the highest-resolution shell are shown in parentheses.
[b]$R_{free} = \sum$Test||Fobs| −|Fcalc||/$\sum$Test |Fobs|, where "Test" is a test set of about 5% of the total reflections randomly chosen and set aside prior to refinement for the complex.
[c]The LSD2-NPAC-H3K4M structure was also determined in the $P2_1$ crystal form, which is similar to the structure in $P3_221$ form (Tables S1 and S2).
[d]Residues Q803 and K75 of LSD2 lie in disallowed regions, and both residues locate at turn regions. The main chain of LSD2 Q803 interacts with FAD, and the side chain forms a hydrogen bond with the residues A546 and S768. Together, these strong interactions lead to a restrained conformation of Q803 of LSD2. A hydrogen-bond interaction is formed between LSD2 residues A74 and G77, which possibly distorts K75 and causes it to lie in a disallowed region.

4B and 4C). The first zinc finger of LSD2 bears little sequence or structural similarity to $Cys_4His_2Cys_2$-type or other types of zinc fingers in the Protein Data Bank. Notably, the Zf-CW domain of LSD2 superimposes with that of ZCWPW1 (Figure S4D), a specific reader of trimethyl H3K4 (Figure S4E)(He et al., 2010). However, the hydrophobic pocket in the LSD2 Zf-CW domain is filled with the side chain of residues L340 and I343 in the adjacent SWIRM domain (Figure S4F). Thus, it is unlikely to interact with other proteins or histone modifications on this surface unless significant conformational change occurs. In LSD2, the two zinc fingers wrap around the SWIRM domain and together form a globular structure that contacts the AO domain. Though the molecular function of these zinc fingers is unclear and warrants future investigation, both are required for LSD2 histone demethylase activity (Fang et al., 2010; Yang et al., 2010). These zinc fingers may play a structural role in stabilizing the conformation of the AO and SWIRM domains. In comparison, the N-terminal region of LSD1 is unstructured and dispensable for LSD1 activity (Shi et al., 2004). Finally, despite some similarities, a significant difference in the LSD2 SWIRM domain is an extended coiled loop connected to the α9-helix, whereas the corresponding region in LSD1 is a short α-helix (Figure 4D). This extended loop and the α9-helix are adjacent to the AO domain, suggesting a possible function in LSD2 enzymatic activity. Indeed, replacing the extended coil loop (residues 273–278) with a flexible peptide sequence (GSGSGS) significantly impaired its enzymatic activity in histone peptide demethylation assays (Figure 4E).

A Dodecapeptide of NPAC Interacts with LSD2

In the cocrystal structures, we unambiguously identified a dodecapeptide of NPAC (residues 214–225) (Figure S2B), whereas other residues of NPAC were not built into the final model due to lack of electron density. This short NPAC peptide binds to LSD2 in a deep hydrophobic groove located between the AO and SWIRM domains, close to its catalytic site (Figure 5A). Specifically, residues H219, F220, L221, and L222 of NPAC are deeply buried in the hydrophobic patch formed by residues L282, V284, L291, L361, F801, and L810 of LSD2 (Figure 5B). Residues 214–217 of NPAC are projected away from this hydrophobic patch and make little contact with LSD2. Notably, LSD2 residues L282, V284, and L291 are located in the α9-helix and in the upstream coiled loop of the LSD2 SWIRM domain, which is one of the structural differences from LSD1 (Figures 5C and S5A), and their conformation is nearly identical to the structure of LSD2 alone (Figures 4D and S3A).

To validate the importance of the dodecapeptide for NPAC-LSD2 interaction, we designed truncation mutants in the NPAC linker region (NP.d7–NP.d9, schematic shown in Figure 5D) and examined their ability to interact with LSD2 (Figure 5E). As expected, deletion of amino acids 214–222 from NP.d3 (NP.d9) completely abolished LSD2 binding (lane 6). NP.d8 (residues 188–252, lane 5), but not NP.d7 (residues 152–186, containing the AT-hook motif, lane 4), was sufficient for LSD2 interaction. Consistent with the results from these binding assays, NP.d8, but not NP.d7 or NP.d9, significantly enhanced LSD2 histone demethylase activity on nucleosomal substrates (Figure 5F). The cocrystal structures predict that residues H219, F220, L221, and L222 of NPAC form the major binding sites for LSD2. Indeed, point mutation and deletion of these critical residues within NPAC 188–252 (NP.M1– NP.M3, partial sequences shown in

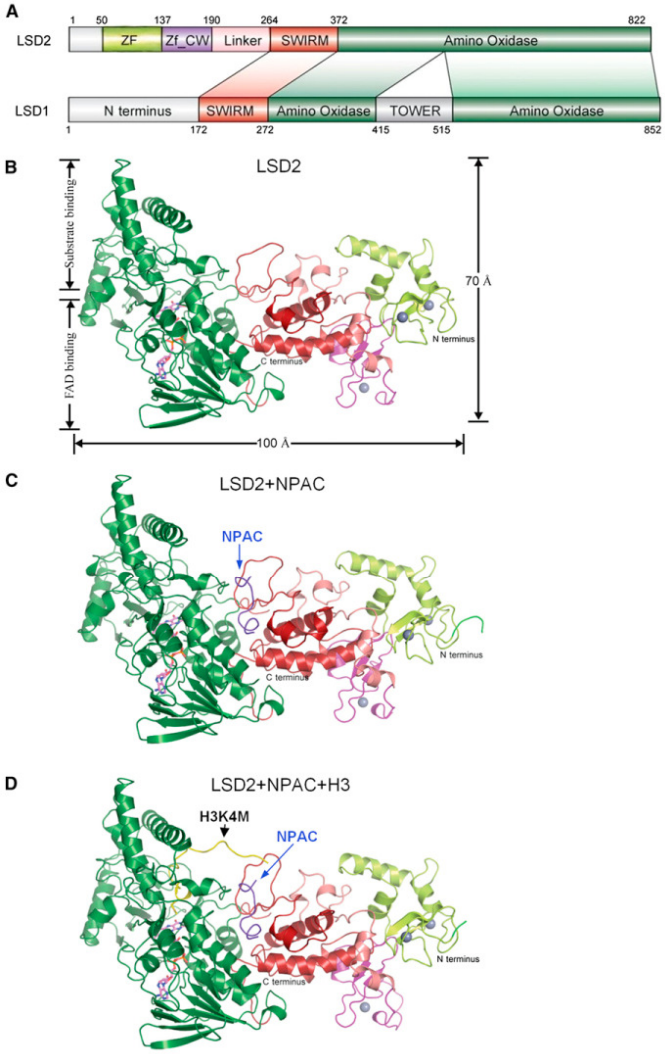


Figure 3. Binding of NPAC Does Not Induce Significant Conformational Changes in LSD2

(A) Comparison of the domain structures of human LSD2 and LSD1. Numbers indicate residue positions at the boundaries of each domain. The ZF domain is shown in lime, the Zf-CW domain in purple, the linker region in pink, the SWIRM domain in red, and the amine oxidase domain in green. The N-terminal flexible regions in both proteins and the tower domain of LSD1 are shown in gray. The same color scheme is used in all structural figures.

(B–D) The crystal structures of LSD2 (B), NPAC-LSD2 heterodimer (C), and the ternary complex of LSD2-NPAC-H3 peptide (D). Disordered regions are shown in dashed lines, and FAD is shown in the stick representation (purple). Three zinc atoms are shown as gray balls. N and C termini of LSD2 are indicated. NPAC and H3K4M peptide are indicated and shown in the ribbon representation in blue and yellow, respectively. See also Figures S2 and S3 and Table S2.

Figure 5G) result in the loss of LSD2 binding (Figure 5H), as well as the loss of the ability to stimulate LSD2 activity (Figure 5I). Collectively, these results confirm the structural predictions and identify the key residues of NPAC whose interaction with LSD2 is important for NPAC cofactor activity.

Interplay among Enzyme, Cofactor, and Substrate

Despite similar interactions observed between the N terminus of the H3 peptide and the catalytic cavity of LSD2 and LSD1 (Figures S4A–S4C), the cocrystal structure of the LSD2-NPAC-H3 peptide complex reveals additional, unique interactions among enzyme, cofactor, and substrate (summarized in Figure S5). A network of hydrogen bonds is formed between the main chains of K18 and L20 of H3K4M peptide and G279, E277, and N276 of LSD2, and between the side chains of H3Q19 and N276 of LSD2 (Figures 6A and 6B). These interactions are unique to LSD2; they are not observed in the cocrystal structure of the ternary complex of LSD1, CoREST, and an H3K4M peptide, in which only residues 1–16 of histone H3 are visible (Forneris et al., 2007). We speculate that the interactions between H3 substrate and LSD2 residues in the extended loop may explain why this region is important for its enzymatic activity (Figure 4E).

The structure of the ternary complex also reveals a unique interaction of H3 peptide with the LSD2-NPAC complex. NPAC F217, together with LSD2 residues Y273, E277, and R285, creates a new hydrophobic patch in the LSD2-NPAC complex that accommodates the side chain of H3L20 (Figure 6C), suggesting a stronger substrate interaction compared to LSD2 alone. In the cocrystal structure, the side chain of H3K18 is in close proximity to D214 and H216 of NPAC, suggesting potential contacts. However, the electron density of the side chain of H3K18 is weak (Figure S2A), suggesting a flexible conformation and weak interactions involving the H3K18 side chain. Taken together, these structural analyses suggest that the dodecapeptide (NPAC

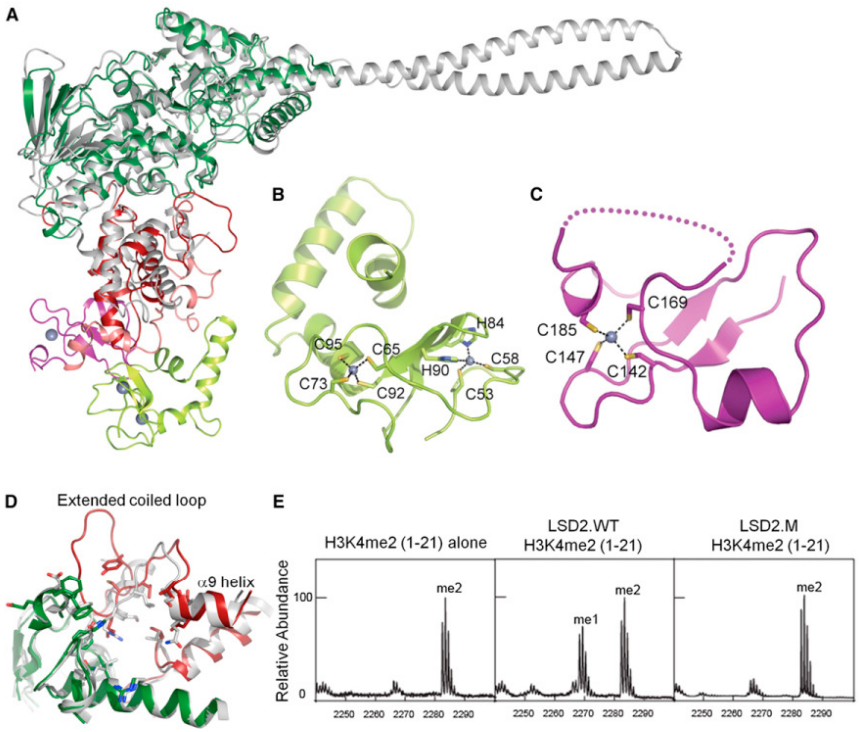


Figure 4. LSD2 Exhibits Both Common and Distinctive Structural Features Compared to LSD1

(A) Structural overlay of LSD2 (colored as in Figure 3) and LSD1 (gray).

(B) Structure of the N-terminal zinc finger of LSD2. Two zinc ions (gray balls) are coordinated with a $Cys_4His_2Cys_2$ motif with the indicated residues shown in stick representation. This zinc-finger domain bears little resemblance to known zinc-finger structures.

(C) Structure of the CW-type zinc finger of LSD2. Residues involved in zinc coordination are labeled and shown in stick representation. The disordered region is indicated by a dotted line.

(D) Superimposed structures of the SWIRM domains of LSD2 and LSD1, colored as in (A). The extended loop and α9-helix in LSD2 SWIRM domain are denoted. The side chains of unconserved residues are presented in stick presentation.

(E) Mutation of the extended loop in LSD2 SWIRM domain impairs its histone demethylase activity. MALDI-TOF MS analyses of the demethylation of H3K4me2 peptides incubated with the indicated proteins are shown. LSD2.WT, wild-type LSD2; LSD2.M, an LSD2 mutant replacing YQPNEC 273–278 with a flexible linker GSGSGS. See also Figure S4 and Table S2.

residues 214–225) is the minimal functional unit for NPAC cofactor activity. The residues H219, F220, L221, and L222 of NPAC form the major binding sites for LSD2 and are responsible for LSD2 and NPAC interaction, whereas NPAC residue F217 probably contributes directly to the cofactor activity of NPAC by stabilizing the enzyme-substrate complex.

F217 of NPAC Stabilizes Enzyme-Substrate Interaction and Is Essential for Cofactor Activity

To determine whether the dodecapeptide (NPAC residues 214–225) is the minimal functional unit for NPAC cofactor activity and whether F217 plays a critical role, we synthesized the wild-type and mutant dodecapeptides and examined their cofactor activities (NP.M4– NP.M6, sequences shown in Figure 6D). As expected, wild-type NPAC peptide is sufficient to enhance LSD2 activity in nucleosome demethylation assays (Figure 6E, comparing lane 3 to lane 2), whereas the F217A single mutation (NP.M6) abolishes cofactor activity (lane 6). In contrast, mutations of D214 and H216 (NP.M4 and NP.M5) have a marginal effect on cofactor activity (comparing lanes 4 and 5 to lane 6), consistent with the structural predictions that neither residue makes important contacts with LSD2 or H3 peptide. These results were further confirmed in demethylation assays using H3K4me2 peptide substrate (residues 1–21) (Figure 6F).

We did not detect NPAC cofactor activity when a shorter H3K4me2 peptide (residues 1–15) was used, even though

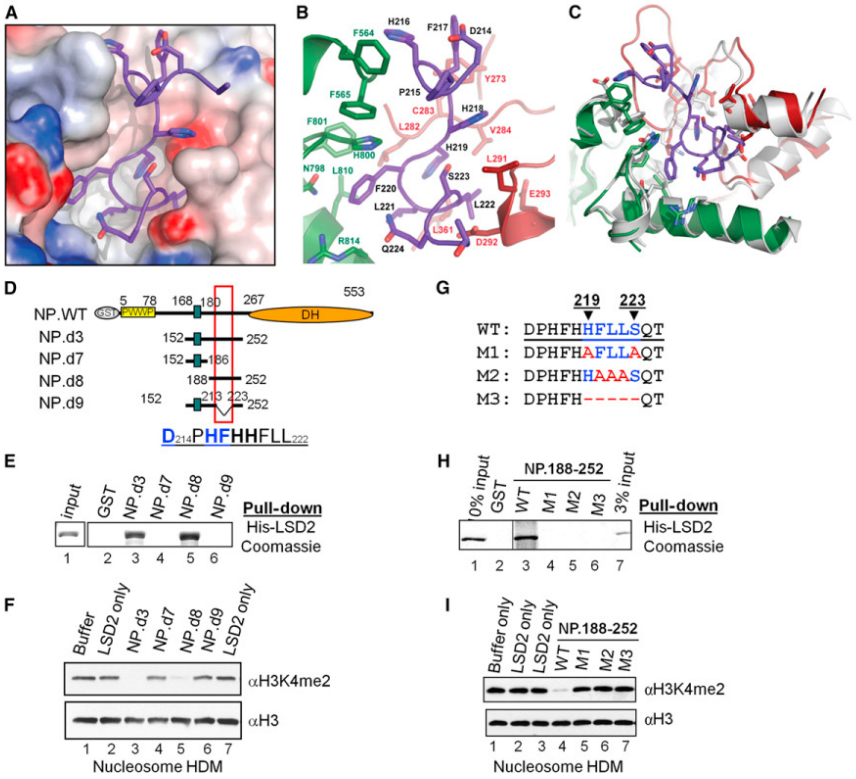


Figure 5. A Dodecapeptide of NPAC Interacts with LSD2

(A and B) Accommodation of NPAC residues in a hydrophobic pocket between the AO and SWIRM domains of LSD2. The surface rendering and ribbon representation of LSD2 are shown in (A) and (B), respectively. Critical residues involved in the interaction are shown in stick representation and colored in green (from the LSD2 AO domain), red (from the LSD2 SWIRM domain), and purple (from NPAC), respectively.

(C) Structural superimposition of the NPAC binding site of LSD2 and the corresponding regions of LSD1. The LSD2 structures are nearly identical with or without NPAC binding.

(D) Schematic of deletion mutants of the NPAC linker region (NP.d7–NP.d9). The red box marks the position of NPAC residues 214–225.

(E) Examination of LSD2 binding of NPAC deletion mutants by GST pull-down. His-LSD2 in pull-down products were separated by SDS-PAGE and visualized by Coomassie blue staining.

(F) Examination of the cofactor activity of NPAC mutants described in (D) using nucleosome demethylation assays. Immunoblots using the indicated antibodies are shown.

(G) Partial sequences of NPAC mutations (NP.M1–NP.M3, residues 188–252) disrupting the potential NPAC binding site for LSD2 interaction.

(H) GST-pull-down analyses of LSD2 binding by NPAC mutants in (G). Coomassie blue staining of His-LSD2 in input and GST-pull-down complexes is shown.

(I) Cofactor activity analyses of NPAC mutants described in (G) using nucleosomal demethylation assays.

weak demethylation activity of this substrate was detected using a high concentration of LSD2 (Figures S6A and S6B). This result supports our model that H3L20 is important for NPAC cofactor activity in making contact with the LSD2-NPAC complex. The significant differences between H3K4me2 peptides 1–15 and 1–21 suggest that the length of the H3 peptide may affect LSD2 enzymatic activity, thus potentially influencing NPAC cofactor activity. To investigate this further, we synthesized a longer H3K4me2 peptide (H3 residues 1–44). Interestingly, the longer peptide seems to be a better substrate for LSD2. Under identical conditions, LSD2 demethylated a large majority of the longer H3K4me2 peptide (1–44), producing H3K4me1 and also H3K4me0 products, whereas only around 30% of H3K4me2 of the short peptide (1–21) was converted to H3K4me1. These observations indicate that H3 residues 16–44 may interact with LSD2 outside its catalytic cavity and that these

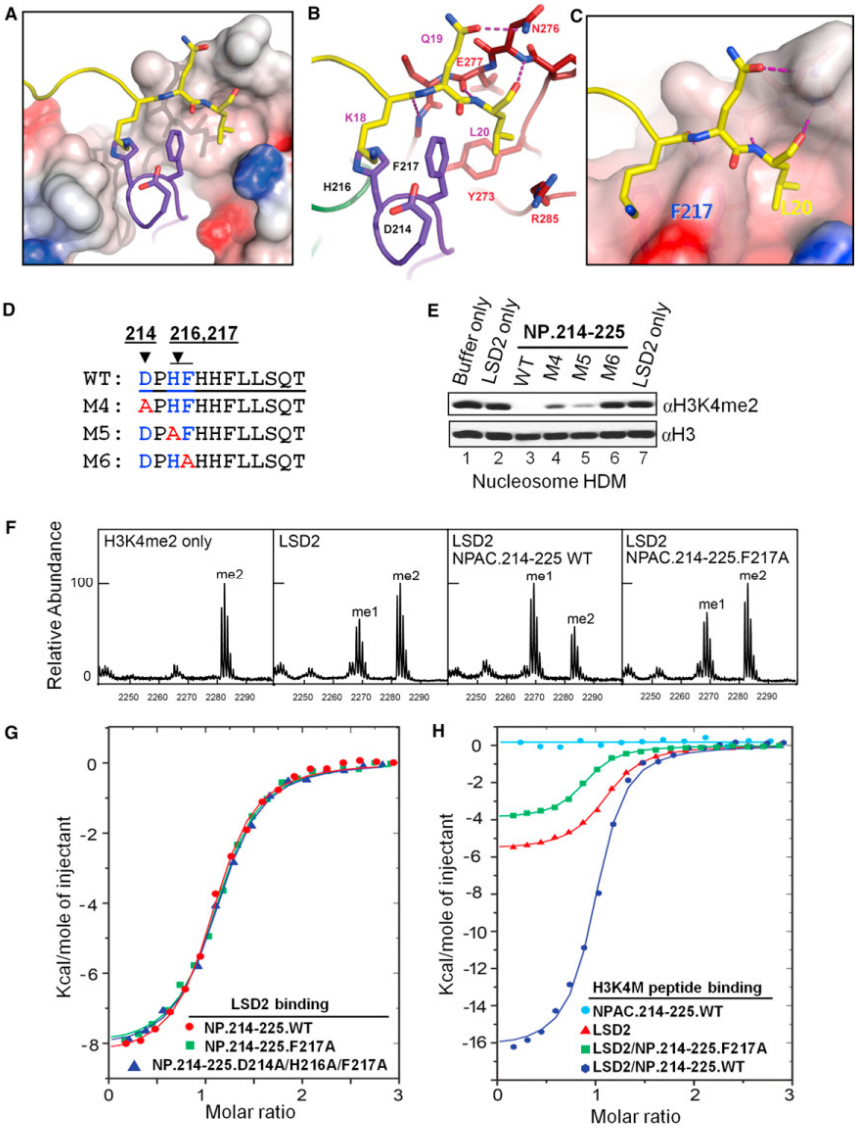


Figure 6. The NPAC Dodecapeptide Stimulates LSD2 Histone Demethylase Activity by Assisting Enzyme-Substrate Interaction

(A and B) A close-up view of the interplay of LSD2, NPAC, and H3K4M peptide in the cocrystal structure. The surface rendering and ribbon representation of LSD2 are shown in (A) and (B), respectively. H3K4M peptide (histone H3 residues 1–21, with Lys4 replaced with a Met) and NPAC (residues 214–225) are shown in the ribbon representation and colored in yellow and purple, respectively. Critical residues involved in H3 peptide interaction are shown in stick representation. Hydrogen bonds are indicated by dashed lines.

(C) H3L20 makes a new contact with the NPAC-LSD2 complex, with its side chain inserted in a hydrophobic patch formed by NPAC F217 and LSD2 residues.

(D) Sequences of wild-type (NP.WT) and mutant NPAC dodecapeptides (NP.M4–NP.M6) disrupting the potential interaction of NPAC with H3K4M peptide in the LSD2-NPAC-H3K4M peptide ternary complex.

(legend continued on next page)

interactions significantly influence H3K4 demethylation efficiency. Notably, the wild-type NPAC protein can robustly stimulate LSD2 demethylation of the longer peptide (1–44), and the cofactor activity of F217A NPAC mutant was very weak or undetectable (Figure S6C).

To further characterize the importance of NPAC F217 for cofactor activity, we compared full-length wild-type and F217A NPAC protein in nucleosome demethylation assays. Using a threshold amount of LSD2 where no obvious nucleosome demethylase activity was observed (Figure S6D, lanes 2 and 9), we did not detect cofactor activity of the F217A mutant, whereas wild-type NPAC showed robust activity (Figure S6D, lanes 10–14). However, residual stimulatory activity of full-length NPAC F217A protein was detected when more NPAC and LSD2 proteins were used (Figure S6Ea, lanes 4 and 9). Wild-type NPAC clearly has a cofactor activity far superior to that of the F217A mutant (Figures S6Db and S6Eb, comparing lanes 3 versus 4 and 8 versus 9). This demonstrates the central role of F217 in NPAC cofactor activity and emphasizes the importance of the interactions of the H3 tail along the surface of the LSD2 enzyme to demethylase activity.

To investigate the functional role of NPAC F217 in its cofactor activity, we first examined the effect of the F217A mutation on LSD2 binding. Isothermal titration calorimetry (ITC) studies showed that wild-type, F217A, and D214A/H216A/F217A triple mutant peptides bind to LSD2 equally well, with a binding affinity (Kd) of 0.92 ± 0.08 μM, 0.93 ± 0.07 μM, and 0.99 ± 0.08 μM, respectively (Figure 6G and Table S3). Thus, the side chains of NPAC residues F217, D214, and H216 are not involved in LSD2 interaction, consistent with the structural predictions. Importantly, the result demonstrates that the inactivation of NPAC F217A mutant is not due to compromised LSD2 interaction.

Next, we investigated the effect of wild-type and mutant NPAC dodecapeptides on LSD2-H3K4M peptide binding via ITC. Shown in Figure 6H, H3K4M peptide had a higher affinity to the LSD2-NPAC complex (blue line) than to LSD2 alone (red line). Importantly, the F217A mutation (NP.M6) significantly diminished the ability of NPAC to stabilize the interaction between LSD2 and the H3K4M peptide (green line). Wild-type NPAC peptide had no appreciable affinity to the H3K4M peptide (cyan line). The thermodynamic features of these interactions are summarized in Table S4. Taken together, these results confirm that NPAC residue F217 directly contributes to the cofactor activity of NPAC by stabilizing the enzyme-substrate complex.

DISCUSSION

The present study identifies NPAC/GLYR1 as a cofactor specific for the LSD2/KDM1b histone demethylase. Our structural and biochemical studies have determined the minimal functional segment of NPAC (a dodecapeptide, residues 214–225) responsible for its cofactor activity. NPAC residues 219–223 interact directly with LSD2. This interaction aligns NPAC residue F217 with the hydrophobic patch on the surface of LSD2, creating a new binding pocket that accommodates the side chain of L20 on histone H3. As a result, NPAC stabilizes the interaction between LSD2 and histone H3 substrates, facilitating H3K4 demethylation. This study thus provides a detailed molecular model precisely illustrating a mechanism of cofactor-assisted histone demethylation.

Histone demethylases may associate with transcriptional factors or chromatin binding proteins and sometimes may themselves contain chromatin-binding modules. All of these present important mechanisms in targeting histone demethylases to specific loci that may consequently increase local enzyme concentration and facilitate histone demethylation. The proposed mechanism for LSD1 cofactor CoREST fits this model (Yang et al., 2006). As a putative H3K36me3 binding protein, NPAC may also play an important role in targeting LSD2 in the human genome. However, independent of its PWWP chromatin-binding module, an NPAC dodecapeptide (residues 214–225) shows robust cofactor activity, stimulating LSD2 demethylation of both nucleosomes and synthetic histone peptides. Thus, we show a cofactor activity independent of the targeting effect, challenging the dogma of histone demethylase regulation.

Our finding answers an intriguing question: What happens beyond the tethering of a histone demethylase to nucleosomes? We propose that the interactions of the H3 tail with LSD2 outside of its catalytic cavity play an important role in regulating histone demethylation efficiency. Structural and biochemical data indicate that NPAC cofactor activity is centered on residue F217 and its ability to assist LSD2 interacting with histone H3L20. The effect of this single interaction on LSD2 activity is striking considering the extensive interactions already in place, particularly between histone H3 residues 1–16 and the LSD2 catalytic domain (Figure S5B). It suggests that even weak interactions of the H3 tail along the surface of the enzyme can be important. Indeed, we observed that LSD2 prefers longer H3K4me2 peptides (Figures S6A–S6C). LSD1 shows no activity to H3K4me1 peptide 1–16, but can demethylate peptides 1–21 and 1–30 with similar efficiency (Forneris et al., 2005). Thus, H3 residues 17–21 are important for LSD1 demethylase activity even though the interaction was observed in the cocrystal structure (Forneris et al., 2007; Yang et al., 2006). We speculate that stabilizing the interaction between the H3 tail and histone demethylases, either by protein factors similar to NPAC or by additional histone modifications on the H3 tail, may present an important mechanism in regulating histone demethylase activity.

(E) NPAC F217 is essential for its cofactor activity. Immunoblots of nucleosome demethylation assays are shown.
(F) The wild-type NPAC dodecapeptide, but not the F217A mutant, can stimulate LSD2-mediated demethylation of H3K4me2 peptides.
(G) Mutations of D214, H216, and F217 of NPAC do not affect LSD2 binding. ITC enthalpy plots of wild-type and mutant NPAC dodecapeptide binding to LSD2 are shown. Indicated NPAC peptides were injected into LSD2 containing cuvettes.
(H) ITC enthalpy plots of the binding of the H3K4M peptide to LSD2 and LSD2 in complex with either wild-type (WT) or F217A (M6) NPAC peptides (residues 214–225). H3K4M peptide (residues 1–21) was injected into cuvettes containing the indicated combination of LSD2 and NPAC peptides. See also Figures S5 and S6 and Tables S3 and S4.

Though CoREST may contribute to cofactor activity for LSD1 through a docking mechanism, there are at least two pieces of evidence indicating additional mechanisms exist. We and others have shown that the CoREST linker region without the two SANT domains retains significant cofactor activity for LSD1 (Lee et al., 2005; Shi et al., 2005). Moreover, Forneris and colleagues showed that CoREST-bound LSD1 exhibits a 2-fold increase in catalytic rate using H3K4me2 peptide substrate (residues 1–21) (Forneris et al., 2007). Our findings presented here with NPAC and LSD2 provide a plausible mechanism to be explored in the context of CoREST and LSD1.

Findings from the present study have significantly advanced our current understanding of cofactor-mediated regulation of histone lysine demethylases in many other aspects. Our findings suggest that the regulation of the enzymatic activity by associated cofactors is probably a general mechanism underlying the regulation of KDM function. Also, the high selectivity of cofactors for histone demethylases, with each cofactor preferentially and specifically regulating its associated KDM, but not the others, may have significant biological implications by defining their specific functional loci in the genome. The present study does not exclude the possibility that one common cofactor may work for several KDMs under certain circumstances; however, it is also possible that each KDM may have more than one cofactor to regulate its diverse functions in distinctive biological processes. For example, LSD1 is regulated by several associating factors besides CoREST, including BHC80 (Lan et al., 2007), MTA-2 (Wang et al., 2009), and nuclear receptors such as estrogen and androgen receptors (Metzger et al., 2005; Nair et al., 2010).

Of particular note, our biochemical and structural analyses suggest that NPAC facilitates H3K4-specific demethylase activity of LSD2 but does not promote switching to H3K9 demethylase activity. Similar to LSD1, the cocrystal structure reveals extensive interactions between the histone H3 tail and the enzyme. The side chain of H3K4M fits nicely in the catalytic site, consistent with the robust H3K4 demethylation activity. In contrast, the side chain of H3K9 is distant from the FAD N5 atom, making LSD2 unfavorable as a potential H3K9me1 or H3K9me2 demethylase (Figures S3B–S3C). However, mouse LSD2 has been reported to possess H3K9 demethylase activity (van Essen et al., 2010). Thus, it remains unclear how LSD2, and arguably also LSD1, might demethylate H3K9. Binding of NPAC does not induce conformational changes significant enough to allow the switch of LSD2 substrate selectivity (H3K4 versus H3K9, Figure S3). However, we do not exclude the possibility that binding of a yet-unidentified cofactor to LSD2 may significantly change its conformation, enabling H3K9 demethylation.

Finally, the findings from the present study have significant biological, clinical, and therapeutic implications. Abrogated expression or enzymatic activity of histone demethylases has been strongly implicated in human diseases such as cancer (Chi et al., 2010; Esteller, 2008; Smith et al., 2007). Therefore, deciphering the regulatory mechanisms of histone demethylases is crucial for understanding their biological and pathophysiological functions (Ng et al., 2007; Chen et al., 2006b; Horton et al., 2010; Shi, 2007). However, this area of research currently remains under investigation. Our study provides a potential foundation for the rational design of specific inhibitors or activators of histone demethylases based on their selective interaction with corresponding cofactors. This is significant, as most current LSD inhibitors effectively inhibit both LSD1 and LSD2 indiscriminately, which is not surprising given their nearly indistinguishable catalytic domains (Binda et al., 2010; Stavropoulos and Hoelz, 2007). Thus, these findings are valuable for future translational research toward epigenetic medicines.

EXPERIMENTAL PROCEDURES

Crystallization and Data Collection

All crystals were grown using the hanging-drop vapor diffusion method. LSD2 (residues 51–822) and NPAC (residues 152–268) were used. The LSD2 used for crystallization of the enzyme alone contains N-terminal extra residues (PLGSEFKGLRRR), whereas the LSD2 used for other crystallization contains extra residues (GPGS) that result from 3C cleavage. The H3 peptide used for crystallization is ARTMQTARKSTGGKAPRKQLA (H3K4M, residues 1–21). Crystals of LSD2 alone were grown in conditions with reservoir containing 7% polyethylene glycol (PEG) 8000, 200 mM NaCl, and 100 mM Na_2HPO_4-KH_2PO_4 (pH 6.0) at 18°C. The LSD2-NPAC complex was grown in conditions with buffer consisting of 10% PEG 3350, 20 mM citric acid, and 30 mM Bis-tris propane at 4°C. The LSD2-NPAC-H3K4M complex was crystallized in two forms. One belongs to space group $P2_1$ in conditions with reservoir containing 10% PEG 3350, 20 mM citric acid, and 30 mM Bis-tris propane, and the other belongs to space group $P3_221$ in conditions with reservoir containing 10% PEG 3350, 100 mM NH_4I, and 100 mM MES (pH 6.2) at 4°C (Table 1 and Table S1). All crystals were slowly equilibrated with a cryoprotectant buffer containing reservoir buffer plus 15% glycerol (v/v) and were flash frozen in a cold nitrogen stream at −173°C. All crystals were examined on an X8 PROTEUM system (Bruker AXS), and data sets were collected on beamline BL17U at the Shanghai Synchrotron Radiation Facility (SSRF, Shanghai, China). All data were processed using the program HKL-2000 (Otwinowski and Minor, 1997).

Structure Determination

The structure of LSD2 alone was determined by molecular replacement using the LSD1 structure (2V1D.PDB) as a searching model (Forneris et al., 2007) in $P2_12_12_1$ form. The crystals contain two molecules in one asymmetric unit. Rotation and translation function searches were performed with the program PHASER (McCoy et al., 2005). The structures of LSD2-NPAC and LSD2-NPAC-H3K4M were determined with the Fourier difference method, and the models were manually built with the Crystallographic Object-Oriented Toolkit (COOT) (Emsley and Cowtan, 2004). All refinements were performed using the refinement module phenix.refine of the PHENIX package (Adams et al., 2002). The model quality was checked with the PROCHECK program (Laskowski et al., 1993), which showed good stereochemistry according to the Ramachandran plot for all structures. The structure-similarity search was performed with the DALI Server (Holm et al., 2008), and structure superimposition was performed with COOT (Emsley and Cowtan, 2004). Even though the $Cys_4His_2Cys_2$ motif of the LSD2-Zf domain (residues 51–137) resembles AN1-type zinc fingers, its structure bears little similarity to AN1-Zfs or other zinc fingers in the PDB. All structure figures were generated with PyMOL (DeLano, 2002). Statistics of the structure determination and refinement are summarized in Table 1 and Table S1.

Histone Demethylase Activity Assays

Histone peptides are purchased from Millipore or custom synthesized. The purity of all peptides is >95%, as determined by high-performance liquid chromatography and MS. In vitro histone demethylase activity assays were performed as described (Shi et al., 2004). In brief, purified LSD2 and NPAC-derived peptides or proteins were incubated with 50 μM H3K4me2 peptides (residues 1–21) in 50 mM Tris-HCl (pH 8.5), 50 mM KCl, 5 mM $MgCl_2$, and 5% glycerol at 37°C for 30 min. The products were desalted through a C18 Zip-tip (Millipore) and analyzed on a MALDI-TOF micro MX mass spectrometer

(ABI 4700, Applied Biosystems). The laser intensity was kept constant for all of the samples. All MS data were processed using Data Explorer 4.5 (Applied Biosystems). Both full-length and truncated LSD2 (residues 51–822) were used for histone peptide demethylation assays, and similar results were obtained.

For nucleosome demethylation assays, typically 0.5 μg of full-length His-LSD2, 1 μg of NPAC protein or peptides, and 2 μg of nucleosomes purified from HeLa cells were used, and demethylation efficiency was analyzed by SDS-PAGE electrophoresis and immunoblot using methylation-specific histone H3 antibodies as previously described (Shi et al., 2004).

ACCESSION NUMBERS

The atomic coordinates of the structures in this work have been deposited in the Protein Data Bank with accession codes 4GU1 for LSD2, 4GUT for LSD2-NPAC, 4GUR for LSD2-NPAC-H3 (1–20; Space Group $p2_1$), and 4GUS for LSD2-NPAC-H3(1–20; Space Group $p3_221$).

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures, four tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2012.11.019.

ACKNOWLEDGMENTS

We thank staff members of beamline BL17U at SSRF for assistance in data collection and staff members of the Biomedical Core Facility, Fudan University for mass spectrometry analyses. We thank James Chou, Paco Kang, and Lisa Schwartz for critical reading of the manuscript. This work was supported by grants from the National Institutes of Health (5R01GM078458) and Brigham and Women's Hospital Funds to Sustain Research Excellence to Y.G.S., and the National Basic Research Program of China (2011CB965300 and 2009CB918600), the National Natural Science Foundation of China (81130047, 31270779, 31030019, 11079016, and 30870493), and the Fok Ying Tung Education Foundation (20090071220012).

Received: April 12, 2012
Revised: September 21, 2012
Accepted: November 16, 2012
Published: December 20, 2012

REFERENCES

Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr. D Biol. Crystallogr. *58*, 1948–1954.

Allis, C.D., Berger, S.L., Cote, J., Dent, S., Jenuwien, T., Kouzarides, T., Pillus, L., Reinberg, D., Shi, Y., Shiekhattar, R., et al. (2007). New nomenclature for chromatin-modifying enzymes. Cell *131*, 633–636.

Bernstein, B.E., Meissner, A., and Lander, E.S. (2007). The mammalian epigenome. Cell *128*, 669–681.

Bhaumik, S.R., Smith, E., and Shilatifard, A. (2007). Covalent modifications of histones during development and disease pathogenesis. Nat. Struct. Mol. Biol. *14*, 1008–1016.

Binda, C., Valente, S., Romanenghi, M., Pilotto, S., Cirilli, R., Karytinos, A., Ciossani, G., Botrugno, O.A., Forneris, F., Tardugno, M., et al. (2010). Biochemical, structural, and biological evaluation of tranylcypromine derivatives as inhibitors of histone demethylases LSD1 and LSD2. J. Am. Chem. Soc. *132*, 6827–6833.

Chen, Y., Yang, Y., Wang, F., Wan, K., Yamane, K., Zhang, Y., and Lei, M. (2006a). Crystal structure of human histone lysine-specific demethylase 1 (LSD1). Proc. Natl. Acad. Sci. USA *103*, 13956–13961.

Chen, Z., Zang, J., Whetstine, J., Hong, X., Davrazou, F., Kutateladze, T.G., Simpson, M., Mao, Q., Pan, C.H., Dai, S., et al. (2006b). Structural insights into histone demethylation by JMJD2 family members. Cell *125*, 691–702.

Chi, P., Allis, C.D., and Wang, G.G. (2010). Covalent histone modifications—miswritten, misinterpreted and mis-erased in human cancers. Nat. Rev. Cancer *10*, 457–469.

Chinnadurai, G. (2007). Transcriptional regulation by C-terminal binding proteins. Int. J. Biochem. Cell Biol. *39*, 1593–1607.

Ciccone, D.N., Su, H., Hevi, S., Gay, F., Lei, H., Bajko, J., Xu, G., Li, E., and Chen, T. (2009). KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. Nature *461*, 415–418.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System (http://www.pymol.org).

Dou, Y., Milne, T.A., Ruthenburg, A.J., Lee, S., Lee, J.W., Verdine, G.L., Allis, C.D., and Roeder, R.G. (2006). Regulation of MLL1 H3K4 methyltransferase activity by its core components. Nat. Struct. Mol. Biol. *13*, 713–719.

Egger, G., Liang, G., Aparicio, A., and Jones, P.A. (2004). Epigenetics in human disease and prospects for epigenetic therapy. Nature *429*, 457–463.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Esteller, M. (2008). Epigenetics in cancer. N. Engl. J. Med. *358*, 1148–1159.

Fang, R., Barbera, A.J., Xu, Y., Rutenberg, M., Leonor, T., Bi, Q., Lan, F., Mei, P., Yuan, G.C., Lian, C., et al. (2010). Human LSD2/KDM1b/AOF1 regulates gene transcription by modulating intragenic H3K4me2 methylation. Mol. Cell *39*, 222–233.

Forneris, F., Binda, C., Vanoni, M.A., Battaglioli, E., and Mattevi, A. (2005). Human histone demethylase LSD1 reads the histone code. J. Biol. Chem. *280*, 41360–41365.

Forneris, F., Binda, C., Adamo, A., Battaglioli, E., and Mattevi, A. (2007). Structural basis of LSD1-CoREST selectivity in histone H3 recognition. J. Biol. Chem. *282*, 20070–20074.

He, F., Umehara, T., Saito, K., Harada, T., Watanabe, S., Yabuki, T., Kigawa, T., Takahashi, M., Kuwasako, K., Tsuda, K., et al. (2010). Structural insight into the zinc finger CW domain as a histone modification reader. Structure *18*, 1127–1139.

Holm, L., Kääriäinen, S., Rosenström, P., and Schenkel, A. (2008). Searching protein structure databases with DaliLite v.3. Bioinformatics *24*, 2780–2781.

Horton, J.R., Upadhyay, A.K., Qi, H.H., Zhang, X., Shi, Y., and Cheng, X. (2010). Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. Nat. Struct. Mol. Biol. *17*, 38–43.

Lan, F., Collins, R.E., De Cegli, R., Alpatov, R., Horton, J.R., Shi, X., Gozani, O., Cheng, X., and Shi, Y. (2007). Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. Nature *448*, 718–722.

Lan, F., Nottke, A.C., and Shi, Y. (2008). Mechanisms involved in the regulation of histone lysine demethylases. Curr. Opin. Cell Biol. *20*, 316–325.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. *26*, 283–291.

Lee, M.G., Wynder, C., Cooch, N., and Shiekhattar, R. (2005). An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. Nature *437*, 432–435.

McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C., and Read, R.J. (2005). Likelihood-enhanced fast translation functions. Acta Crystallogr. D Biol. Crystallogr. *61*, 458–464.

Metzger, E., Wissmann, M., Yin, N., Müller, J.M., Schneider, R., Peters, A.H., Günther, T., Buettner, R., and Schüle, R. (2005). LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. Nature *437*, 436–439.

Nair, S.S., Nair, B.C., Cortez, V., Chakravarty, D., Metzger, E., Schüle, R., Brann, D.W., Tekmal, R.R., and Vadlamudi, R.K. (2010). PELP1 is a reader of histone H3 methylation that facilitates oestrogen receptor-alpha target gene

activation by regulating lysine demethylase 1 specificity. EMBO Rep. *11*, 438–444.

Ng, S.S., Kavanagh, K.L., McDonough, M.A., Butler, D., Pilka, E.S., Lienard, B.M., Bray, J.E., Savitsky, P., Gileadi, O., von Delft, F., et al. (2007). Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. Nature *448*, 87–91.

Nottke, A., Colaiácovo, M.P., and Shi, Y. (2009). Developmental roles of the histone lysine demethylases. Development *136*, 879–889.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. In Methods in Enzymology, Volume 276: Macromolecular Crystallography, part A, C.W. Carter, Jr. and R.M. Sweet, eds. (New York: Academic Press), pp. 307–326.

Rice, J.C., and Allis, C.D. (2001). Histone methylation versus histone acetylation: new insights into epigenetic regulation. Curr. Opin. Cell Biol. *13*, 263–273.

Ruthenburg, A.J., Allis, C.D., and Wysocka, J. (2007). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. Mol. Cell *25*, 15–30.

Shi, Y. (2007). Histone lysine demethylases: emerging roles in development, physiology and disease. Nat. Rev. Genet. *8*, 829–833.

Shi, Y., Sawada, J., Sui, G., Affar, B., Whetstine, J.R., Lan, F., Ogawa, H., Luke, M.P., Nakatani, Y., and Shi, Y. (2003). Coordinated histone modifications mediated by a CtBP co-repressor complex. Nature *422*, 735–738.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941–953.

Shi, Y.J., Matson, C., Lan, F., Iwase, S., Baba, T., and Shi, Y. (2005). Regulation of LSD1 histone demethylase activity by its associated factors. Mol. Cell *19*, 857–864.

Smith, L.T., Otterson, G.A., and Plass, C. (2007). Unraveling the epigenetic code of cancer for therapy. Trends Genet. *23*, 449–456.

Southall, S.M., Wong, P.S., Odho, Z., Roe, S.M., and Wilson, J.R. (2009). Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. Mol. Cell *33*, 181–191.

Stavropoulos, P., and Hoelz, A. (2007). Lysine-specific demethylase 1 as a potential therapeutic target. Expert Opin. Ther. Targets *11*, 809–820.

Stavropoulos, P., Blobel, G., and Hoelz, A. (2006). Crystal structure and mechanism of human lysine-specific demethylase-1. Nat. Struct. Mol. Biol. *13*, 626–632.

Tahiliani, M., Mei, P., Fang, R., Leonor, T., Rutenberg, M., Shimizu, F., Li, J., Rao, A., and Shi, Y. (2007). The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation. Nature *447*, 601–605.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature *439*, 811–816.

van Essen, D., Zhu, Y., and Saccani, S. (2010). A feed-forward circuit controlling inducible NF-κB target gene activation by promoter histone demethylation. Mol. Cell *39*, 750–760.

Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G., and Mann, M. (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell *142*, 967–980.

Wang, Y., Zhang, H., Chen, Y., Sun, Y., Yang, F., Yu, W., Liang, J., Sun, L., Yang, X., Shi, L., et al. (2009). LSD1 is a subunit of the NuRD complex and targets the metastasis programs in breast cancer. Cell *138*, 660–672.

Wilson, J.R. (2007). Targeting the JMJD2A histone lysine demethylase. Nat. Struct. Mol. Biol. *14*, 682–684.

Yang, M., Gocke, C.B., Luo, X., Borek, D., Tomchick, D.R., Machius, M., Otwinowski, Z., and Yu, H. (2006). Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. Mol. Cell *23*, 377–387.

Yang, Z., Jiang, J., Stewart, D.M., Qi, S., Yamane, K., Li, J., Zhang, Y., and Wong, J. (2010). AOF1 is a histone H3K4 demethylase possessing demethylase activity-independent repression function. Cell Res. *20*, 276–287.

Appendix B

Cell

Tet3 CXXC Domain and Dioxygenase Activity Cooperatively Regulate Key Genes for *Xenopus* Eye and Neural Development

Yufei Xu,[1,11] Chao Xu,[2,11] Akiko Kato,[3,11] Wolfram Tempel,[2] Jose Garcia Abreu,[4] Chuanbing Bian,[2] Yeguang Hu,[1] Di Hu,[1,5] Bin Zhao,[5] Tanja Cerovina,[2] Jianbo Diao,[5] Feizhen Wu,[5] Housheng Hansen He,[6] Qingyan Cui,[7] Erin Clark,[1] Chun Ma,[1,5] Andrew Barbara,[1] Gert Jan C. Veenstra,[8] Guoliang Xu,[7] Ursula B. Kaiser,[1] X. Shirley Liu,[6,9] Stephen P. Sugrue,[10] Xi He,[4] Jinrong Min,[2,*] Yoichi Kato,[3,*] and Yujiang Geno Shi[1,*]

[1]Division of Endocrinology, Diabetes and Hypertension, Departments of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA
[2]Structural Genomics Consortium and Department of Physiology, University of Toronto, Toronto, ON M5G 1L7, Canada
[3]Department of Biomedical Sciences, Florida State University College of Medicine, Tallahassee, FL 32306, USA
[4]F.M. Kirby Neurobiology Center, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA
[5]Laboratory of Epigenetics, Institutes of Biomedical Science, Fudan University, Shanghai 200032, P.R. China
[6]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, MA 02115, USA
[7]The State Key Laboratory of Molecular Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, P.R. China
[8]Radboud University Nijmegen, Nijmegen Center for Molecular Life Sciences, Nijmegen 6500 HB, The Netherlands
[9]Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[10]Department of Anatomy and Cell Biology, University of Florida, Gainesville, FL 32610, USA
[11]These authors contributed equally to this work
*Correspondence: jr.min@utoronto.ca (J.M.), yoichi.kato@med.fsu.edu (Y.K.), yujiang_shi@hms.harvard.edu (Y.G.S.)
http://dx.doi.org/10.1016/j.cell.2012.11.014

SUMMARY

Ten-Eleven Translocation (Tet) family of dioxygenases dynamically regulates DNA methylation and has been implicated in cell lineage differentiation and oncogenesis. Yet their functions and mechanisms of action in gene regulation and embryonic development are largely unknown. Here, we report that *Xenopus* Tet3 plays an essential role in early eye and neural development by directly regulating a set of key developmental genes. Tet3 is an active 5mC hydroxylase regulating the 5mC/5hmC status at target gene promoters. Biochemical and structural studies further demonstrate that the Tet3 CXXC domain is critical for specific Tet3 targeting. Finally, we show that the enzymatic activity and CXXC domain are both crucial for Tet3's biological function. Together, these findings define Tet3 as a transcription regulator and reveal a molecular mechanism by which the 5mC hydroxylase and DNA binding activities of Tet3 cooperate to control target gene expression and embryonic development.

INTRODUCTION

The process of vertebrate development is established through the integration of several molecular pathways controlled by key regulatory genes and complex epigenetic markings. DNA methylation at the 5-position of cytosine (5mC) is a key epigenetic mark playing crucial roles in vertebrate development (Bestor and Coxon, 1993; Bird, 1986; Reik et al., 2001). Recent studies have demonstrated that the Tet family of 5mC hydroxylases can catalyze the conversion of 5mC to 5-hydroxymethylcytosine (5hmC) (Tahiliani et al., 2009) and further to 5-formylcytosine (5fC) and 5-carboxylcytosine (5CaC) (He et al., 2011; Ito et al., 2011). These studies also suggest that additional modification of 5mC modulated by Tet enzymes may regulate the dynamics of 5mC and its mediated gene regulation (Branco et al., 2012).

The mammalian Tet family has three members, Tet1, Tet2, and Tet3. It has been suggested that both Tet1 and Tet2 play important roles in ES cell lineage specification (Ito et al., 2010; Koh et al., 2011) and that Tet1 regulates DNA methylation and gene expression in mouse ES cells (Ficz et al., 2011; Williams et al., 2011; Wu et al., 2011; Xu et al., 2011b). Mutational inactivation of *TET2* has been reported to associate with decreased 5hmC levels in various myeloid leukemias (Delhommeau et al., 2009; Langemeijer et al., 2009), and Tet2 deficiency leads to increased hematopoietic stem cell self-renewal and myeloid transformation in mouse (Moran-Crusio et al., 2011; Quivoron et al., 2011). Recently, we and others also show that TET1 and TET2 play critical roles in other human cancers, such as melanoma and breast cancer (Hsu et al., 2012; Lian et al., 2012). In addition, Tet3 is the only Tet family member highly expressed in mouse oocytes and zygotes and is responsible for the hydroxylation of 5mC that occurs in the paternal pronucleus of advanced pronuclear-stage zygotes (Gu et al., 2011; Iqbal et al., 2011;

Wossidlo et al., 2011). Conditional knockout of Tet3 in mouse oocytes prevents resetting of DNA methylation patterns in zygotes and impairs reprogramming of transferred somatic nuclei (Gu et al., 2011). Nevertheless, *Tet3*$^{-/-}$ knockout mice are viable through development, but die on postnatal day one (Gu et al., 2011). Taken together, although the discovery of the Tet family of 5mC hydroxylases provides a potential mechanism for the dynamic regulation of DNA methylation, it remains unclear how Tet proteins are recruited to and regulate the expression of target genes, thereby providing linkage to their specific functions in early vertebrate embryonic development.

Although all Tet family members contain a conserved C-terminal catalytic domain, only Tet1 and Tet3 contain the CXXC domain, a potential DNA binding module characterized by two CXXCXXC repeats. The CXXC domains, found in other proteins such as DNMT1, MLL, and CFP1, have been shown to specifically bind to unmethylated CpG dinucleotides and participate in gene transcription regulation (Allen et al., 2006; Pradhan et al., 2008; Xu et al., 2011a). Although our previous study has suggested an important role of the CXXC domain in targeting Tet1 enzyme to specific genomic regions in ES cells (Xu et al., 2011b), the molecular mechanism and biological importance of this domain in Tet1- and Tet3-mediated transcriptional regulation of target genes remain largely unknown.

In this report, we characterize the molecular and biochemical properties and the biological function of Tet3 by using *Xenopus* as a model. Our study shows that Tet3 is essential for early eye and neural development in *Xenopus*. We also demonstrate that several master control genes essential for eye and neural development are Tet3's direct targets, mechanistically linking Tet3 function in transcriptional regulation of these key genes to the developmental phenotypes caused by Tet3 depletion. Using structural and mutational analyses and functional rescue approaches, we show that Tet3's 5mC hydroxylase and the CXXC domain-mediated DNA binding activities cooperate to regulate target gene expression during eye and neural development.

RESULTS

Identification and Characterization of *Xenopus Tet3* Gene Reveal that *Tet3* Is Essential for Early Eye and Neural Development

To understand the biological function of Tet proteins in early embryonic development, we investigated Tet family members in *Xenopus*. Database searches reveal two *Tet* orthologs in *Xenopus tropicalis*, *Tet2* and *Tet3*, but not *Tet1* (Figure S1A available online). We cloned *Xenopus laevis Tet3* (*xlTet3*) gene and subsequently discovered two *xlTet3* isoforms (HQ220207-*xlTet3a* and HQ220208-*xlTet3b*) that exhibit greater than 90% amino acid similarity to *xtTet3* (Figure S1A). Surprisingly, despite extensive searching, we were unable to identify a *Tet1* ortholog in either *X. tropicalis* or *X. laevis*. We, therefore, conclude that the *Xenopus* genome contains only two *Tet*-related genes, *Tet2* and *Tet3*, and lacks a *Tet1* gene. Sequence analysis reveals that similar to mammalian Tet3 proteins (which we cloned, validated, and deposited into GeneBank; HQ220209, human TET3; and HQ423151, mouse Tet3), *Xenopus* Tet3 contains a CXXC domain, a cysteine-rich domain and a double-stranded β helix (DSBH)-containing dioxygenase domain (Figure S1A).

We next examined the expression profile of *xlTet3* during embryogenesis. The temporal expression pattern of *xlTet3* by RT-qPCR reveals that, unlike the high level of *Tet3* mRNA observed in mouse oocytes (Gu et al., 2011; Iqbal et al., 2011; Wossidlo et al., 2011), the *xlTet3* mRNA level in oocytes is very low. *xlTet3* mRNA is also barely detectable at stage 2 (2-cell stage) through stage 7, but increases dramatically from stage 9 (immediately after the initiation of zygotic transcription) to stage 19 and then drops at stage 25 (Figure S1B). The spatial expression pattern of *xlTet3* by in situ hybridization shows that *xlTet3* mRNA is barely detectable at stage 9 (Figure 1A), but clearly detected by stage 14 in the neural plate (Figure 1B). In embryo sections, we also detect *xlTet3* mRNA in the neural plate and notochord (Figures S1C and S1D). Moreover, *xlTet3* mRNA level remains high in the neural tube at stage 19 (Figure 1C) and in the region of the developing brain, eye, branchial arches, cement gland, and spinal cord at stage 25 (Figures 1D and 1E).

To address Tet3 function during embryogenesis, we performed loss-of-function studies by depleting endogenous xlTet3 protein by using the Morpholino antisense oligo (MO) strategy. We first confirmed that *xlTet3* MOs efficiently deplete Tet3 protein (Figure 1F). Control MO or *xlTet3* MOs were injected into two dorsal blastomeres of 4-cell stage embryos. At stage 35, we observe striking developmental abnormalities in 96% (194/201) of *xlTet3* MOs-injected embryos, including malformation of the eye (eyeless), small head, and missing pigmentation along the lateral body, whereas control embryos develop normally (Figure 1G). The Tet3-depleted embryos die between stages 35 and 40. Importantly, these phenotypes are rescued by coinjecting the *xlTet3* MOs-resistant *xtTet3* mRNA (Figure 1G, right). Taken together, these data suggest that Tet3 plays an essential role in early embryogenesis, especially in early eye and neural development.

Tet3 Directly Regulates a Set of Genes Critical for Eye and Neural Development

To gain insight into the underlying molecular mechanisms linking Tet3 function to early eye and neural development, we examined the effect of Tet3 depletion on the expression of a set of key developmental genes, including *pax6* (eye and neural marker), *rx* and *six3* (eye markers), *sox2* (pan-neural marker), *otx2* (anterior neural marker), *sox9* and *snail* (neural crest markers), neurogenin related 1 (*ngn2*) and n-tubulin (*tubb2b*) (primary neuron markers), and *shh* and *ptc-1* (sonic hedgehog signaling) in stage 14 embryos. MOs were injected into one dorsal blastomere of 4-cell stage embryos so that the uninjected side could be used as an internal control. For *shh* and *ptc-1* measurement, we injected MOs into two dorsal blastomeres of 4-cell stage embryos because the expression of *shh* and *ptc-1* is in the midline. As shown by in situ hybridization, the expression of master eye developmental genes, *pax6*, *rx*, and *six3*, is greatly reduced on the *xlTet3* MOs-injected side compared to the uninjected side at stage 14 (Figure 2A). Reduced expression of these eye genes is also seen in the potential eye field on the Tet3-depleted side at stage 19 (Figure S2A). Depletion of Tet3 also inhibits the expression of two primary neuron markers, *ngn2* and *tubb2b*, and two

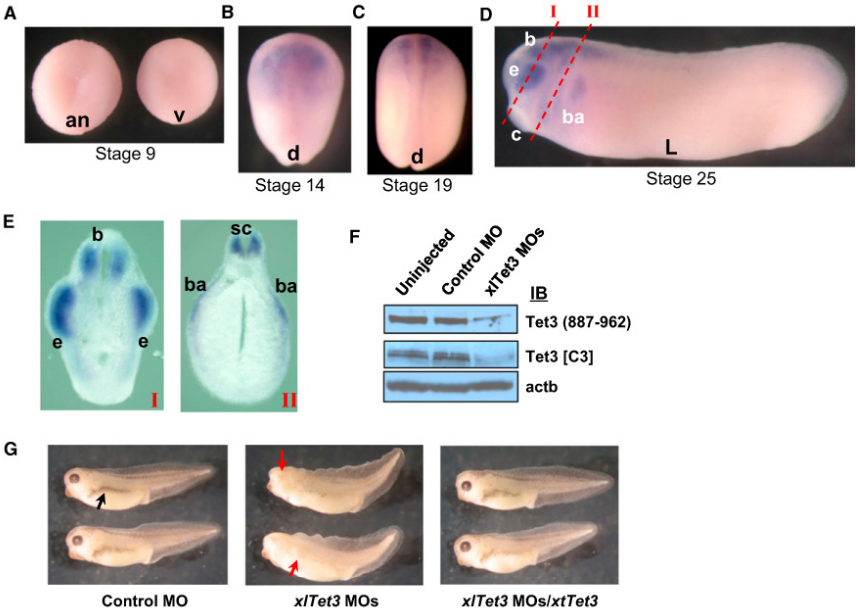


Figure 1. Tet3 Is Important for Early Eye and Neural Development

(A–E) Spatial expression profile of *xlTet3* by in situ hybridization at stage 9 (A), 14 (B), 19 (C), and 25 (D and E). The sites of sections I and II in (E) are noted by red dashed lines in (D). Animal view (an), vegetal view (v), dorsal view (d), lateral view (L), brain (b), eye (e), cement gland (c), branchial arches (ba), and spinal cord (sc).

(F) Western blot showing depletion of endogenous Tet3 protein by *xlTet3* MOs in stage 14 embryos.

(G) Developmental defects in stage 35 embryos caused by Tet3 depletion. Small head, eyeless, and missing pigmentation phenotypes in *xlTet3* MOs-injected embryos are noted by red arrows, and the normal pigmentation in control embryos is noted by a black arrow.

See also Figure S1.

neural crest markers, *sox9* and *snail*, supporting a critical role of Tet3 in neural and neural crest development (Figure 2A). The expression of two major shh signaling components, *shh* and *ptc-1,* is also abolished in Tet3-depleted embryos at stage 14 (Figure 2A), whereas the expression of *otx2* and *sox2* shows no significant changes (Figure S2B). Importantly, the diminished expression of all affected genes is rescued by *xtTet3* mRNA coinjection (Figure 2A), demonstrating the specific regulation of these genes by Tet3. Results by in situ hybridization are further independently confirmed by RT-qPCR assays (Figures 2B and S2C).

To address whether Tet3 directly regulates these key genes, we performed chromatin immunoprecipitation (ChIP) assays. We employed an epitope-tagged ChIP assay by using Flag antibody and *flag-xtTet3* mRNA-injected embryos to explore Tet3 occupancy at the promoters of these genes. We detect strong Tet3 binding at the promoters of *pax6*, *rx*, *six3*, *ptc-1, ptc-2, sox9*, and *ngn2* but not at the promoters of control genes, myosin light chain 2 (*myl2*), and cardiac actin (*actc*) (Figure 2C).

Finally, we tested whether ectopic expression of any one of these downstream Tet3 target genes can rescue the phenotypes caused by Tet3 depletion. We examined the rescue effects of *pax6*, *rx*, or *shh* overexpression. As shown in Figure S2D, none of these can rescue the phenotypes, suggesting that it may be the cumulative effect of the altered expression of a group of target genes, rather than a single gene, that causes the observed developmental defects. Nevertheless, our data strongly suggest that Tet3 is an upstream transcriptional regulator specifically and directly controlling a set of key genes important for early eye and neural development.

Tet3 Is an Active 5mC Hydroxylase that Regulates the 5mC/5hmC Status at Target Gene Promoters

As DNA methylation at gene promoters has regulatory effects on the expression of the associated genes, we asked whether *Xenopus* Tet3 modulates the 5mC/5hmC status at its target gene promoters. To address this question, we first demonstrate that *Xenopus* Tet3 has conserved 5mC hydroxylase activity as mammalian Tet proteins. The xlTet3 catalytic domain (CD) can convert 5mC to 5hmC in cells and in vitro (Figures 3A and 3B). Utilizing two independent methods, anti-5hmC antibody-based dot-blot and T4 Phage β-glucosyltransferase-mediated 5hmC glucosylation assays, we demonstrate that 5hmC exists in the

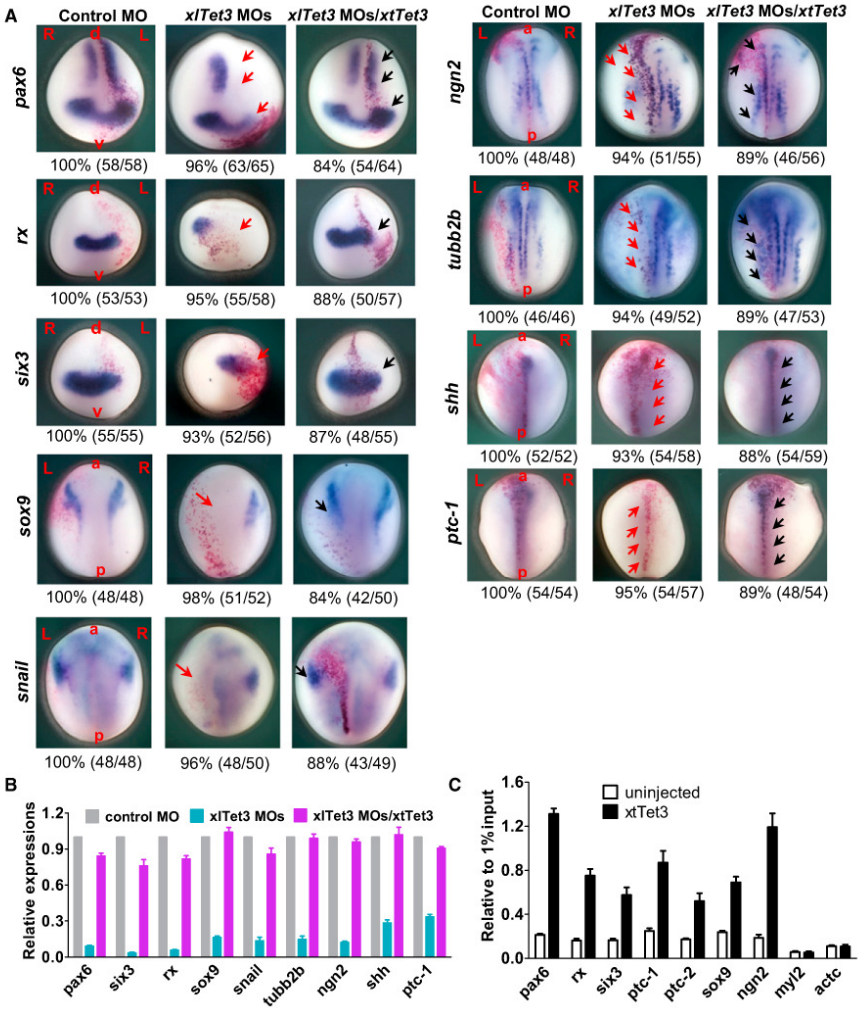


Figure 2. Tet3 Directly Regulates Key Developmental Genes

(A) Expression level changes of developmental genes resulting from Tet3 depletion as shown by in situ hybridization at stage 14. Red arrows show the inhibited expression by *xlTet3* MOs injection, whereas black arrows show rescued expression by *xtTet3* coinjection. L, left; R, right; a, anterior; p, posterior; d, dorsal; v, ventral. The red dots are injection tracer by *β-gal* staining.

(B) RT-qPCR confirms the differentially expressed genes after Tet3 depletion at stage 14. Control MO, *xlTet3* MOs or *xlTet3* MOs/*xtTet3* was injected into two dorsal blastomeres at 4-cell stage. Relative gene expression was normalized to *odc*. Results are shown as mean ± SD (n = 3).

(C) Tet3 occupancy at target gene promoters by ChIP-qPCR assay. Data are presented as mean ± SEM (n = 3).

See also Figure S2.

genome of *Xenopus* embryos and is globally reduced after Tet3 depletion (Figure 3C). However, consistent with undetectable or extremely low levels of 5fC and 5CaC in ES cells and various tissues (Globisch et al., 2010; Ito et al., 2011), we failed to detect 5fC and 5CaC in *Xenopus* embryos (data not shown). Collectively, these data suggest that Tet3 possesses intrinsic 5mC

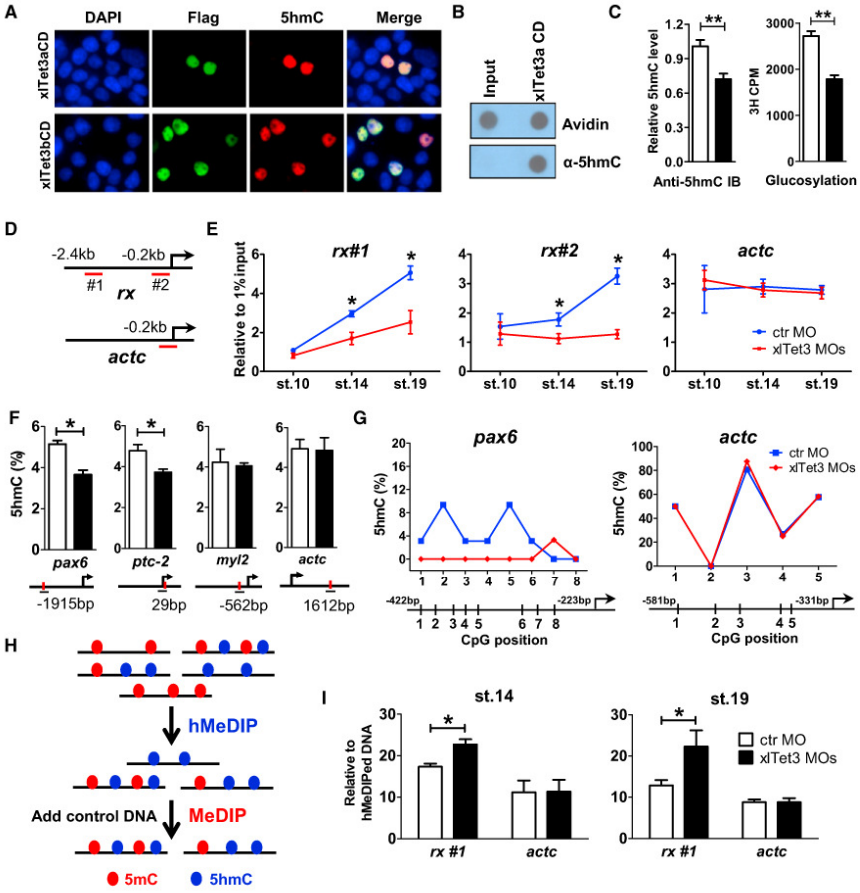


Figure 3. Tet3 Is an Active 5mC Hydroxylase Regulating the 5mC/5hmC Status at Target Gene Promoters

(A) xITet3 CD is sufficient to convert 5mC to 5hmC in HEK293T cells by immunofluorescence analysis. Flag-tagged xITet3 CD protein was detected using Flag antibody.

(B) xITet3a CD protein converts 5mC to 5hmC in vitro by dot-blot assay. Avidin-HRP is used to detect total biotin-labeled DNA, showing equal loading.

(C) Tet3 depletion results in globally decreased 5hmC in stage 14 embryos by dot-blot (left) and 5hmC glucosylation (right) assays. Open bar, control MO; filled bar, *xlTet3* MOs. Data are presented as mean ± SD (n = 3). **p < 0.01.

(D and E) hMeDIP-qPCR to detect dynamic 5hmC level changes in stage 10, 14, and 19 embryos. The targeting region for each primer set is underlined in (D). Results are shown as mean ± SD (n = 3) in (E). *p < 0.05.

(F) Site-specific 5hmC level changes by Tet3 depletion in stage 14 embryos using the *EpiMark* 5mC/5hmC analysis kit. Open bar, control MO; filled bar, *xlTet3* MOs. Red dot indicates MspI/HapII recognition site and each PCR amplified region is underlined. Arrow denotes promoter orientation. Data are shown as mean ± SD (n = 3). *p < 0.05.

(G) TAB-seq analyses of 5hmC status at the promoter of *pax6* (left) and *actc* (right) in stage 14 embryos. The average percent at each CpG site is derived from sequencing of 30–32 clones for *pax6* promoter and 24–26 clones for *actc* promoter.

(H) Schematic diagram of hMeDIP-MeDIP strategy.

(I) hMeDIP-MeDIP qPCR to detect 5mC level changes after Tet3 depletion in stage 14 and 19 embryos. Data are shown as mean ± SD (n = 3). *p < 0.05. The targeting region for each primer set is shown in (D).

See also Figure S3.

hydroxylase activity and is, at least in part, responsible for modulating 5hmC levels in *Xenopus* embryos.

We next asked whether Tet3 regulates 5hmC levels at target gene promoters. We first examined the dynamic 5hmC level changes by hydroxymethylated DNA immunoprecipitation (hMeDIP)-qPCR (Xu et al., 2011b). Consistent with the increasing *Tet3* expression from stage 10 to 19 (Figure S1B), we observe increasing 5hmC levels at a Tet3 target gene promoter, *rx* promoter, but not at the nontargeted *actc* promoter (Figures 3D and 3E). Importantly, Tet3 depletion results in a significant reduction in 5hmC at *rx* promoter at stages 14 and 19, thereby significantly abolishing the increase in 5hmC from stage 10 to 14 to 19 ($p < 0.05$), whereas no 5hmC level changes are observed at *actc* promoter (Figure 3E). However, we failed to detect 5hmC level changes at other Tet3 target gene promoters including *pax6*, *six3*, *sox9*, and *ptc-2* by hMeDIP-qPCR perhaps due to the low 5hmC levels at those promoters and the detection limits of hMeDIP-qPCR. Thus, we applied a more sensitive CCGG site-dependent strategy, *EpiMark* method (Ficz et al., 2011). As exemplified by *pax6* and *ptc-2*, depletion of Tet3 results in a significant reduction in 5hmC at CCGG sites of both promoters ($p < 0.05$), whereas no 5hmC changes are observed at the nontargeted *myl2* and *actc* genes (Figure 3F). To next determine the 5hmC status at multiple CpG sites, we employed the newly developed Tet-assisted bisulfite sequencing (TAB-Seq) approach (Yu et al., 2012), the most sensitive method to specifically detect 5hmC at base-pair resolution. We first confirm that more than 90% of 5mCs in the internal control DNA oligo have been converted by TET1 CD (Figures S3A–S3C). Using *pax6* as an example, we clearly observe a significant decrease in 5hmC at specific CpG sites within the *pax6* promoter after Tet3 depletion ($p < 0.05$ by one-way ANOVA), whereas the 5hmC levels at *actc* promoter are not changed (Figure 3G), consistent with that observed via the *EpiMark* method. Taken together, by employing multiple approaches to analyze changes in 5hmC at promoters after Tet3 depletion, we demonstrate that Tet3 regulates the 5hmC status at its target gene promoters.

Finally, we addressed the extent to which 5hmC level decrease by Tet3 depletion results in a subsequent 5mC level increase. Notably, examining the alteration of 5mC directly resulting from the change of 5hmC at Tet3 target gene promoters is extremely technically challenging. Like most *Xenopus* genes, *Tet3* exhibits specific spatial-temporal expression in embryos (Figures 1B–1E). Thus, changes in 5mC caused by Tet3 depletion are expected to occur only in the small fraction of the embryo expressing Tet3, whereas 5mC remains unchanged in the rest of embryo, effectively diluting the signal from Tet3-expressing cells. Indeed, using the whole embryos, we attempted and failed to detect 5mC changes after Tet3 depletion by multiple approaches, such as methylated DNA immunoprecipitation (MeDIP) and methylated CpG island recovery assay. We therefore developed a targeting strategy, employing hMeDIP-MeDIP qPCR to separate the signal from the noise and examine the 5mC level changes directly resulting from 5hmC level changes by Tet3 depletion (Figure 3H). We first performed hMeDIP to enrich 5hmC-containing genomic DNA fragments that presumably are from Tet-expressing cells. Given that Tet3 depletion results in a 5hmC level decrease but not complete removal, Tet3-targeted DNAs in *xlTet3* MOs-treated embryos still contain 5hmC and can be enriched, albeit with a lower amount than the control embryos. We then carried out MeDIP by using the hMeDIPed DNA as input to examine relative 5mC levels in those 5hmC-containing DNAs given the dynamic conversion of 5mC to 5hmC and the coexistence of 5mC and 5hmC in the same region (Ficz et al., 2011; Yu et al., 2012). We first confirm that the different amount of starting DNA (hMeDIPed DNA) will not introduce a MeDIP efficiency bias, as validated by the equal MeDIP enrichment of the human *NBR2* promoter from control DNA spiked in the hMeDIPed DNA from control MO- or Tet3 MOs-injected embryos (Figure S3D). For proof-of-principle, we examined the *rx* gene promoter because it has the most dramatic changes in 5hmC after Tet3 depletion and can be easily detected (Figure 3E). Indeed, we clearly observe a significant 5mC level increase at *rx* gene promoter after Tet3 depletion by this hMeDIP-MeDIP qPCR method in stage 14 and 19 embryos (Figure 3I), whereas the 5mC level at *actc* promoter is not changed (Figure 3I), validating the reliability of hMeDIP-MeDIP qPCR procedure. Thus, we conclude that Tet3 regulates target gene expression, at least partially, through control of 5mC/5hmC status at the promoter of target genes.

The Tet3 CXXC Domain Possesses Unique DNA Binding Properties

In addition to the dioxygenase domain that confers 5mC hydroxylase activity, Tet3 also contains a potential DNA binding domain, the CXXC domain. In general, the selective DNA binding activity of a transcription factor serves as a key mechanism for action of the transcription factor in gene transcriptional regulation. Sequence alignment indicates that Tet CXXC domains exhibit a conserved overall structure with other CXXC domains (Figure S4A). However, Tet CXXC domains lack a short sequence motif (KFGG) (Figure S4A), which has been shown to be important for the DNA binding activity of the MLL and CFP1 CXXC domains (Allen et al., 2006; Xu et al., 2011a). To address whether the Tet3 CXXC domain may possess unique DNA binding features, we examined the DNA binding ability and specificity of the *Xenopus* Tet3 CXXC domain (aa 58–111) by isothermal titration calorimetry (ITC) assays. The Tet3 CXXC domain strongly binds to various C/G-rich DNA oligos but exhibits virtually no binding activity to the A/T-only DNA oligo (Figures 4A–4E and Table S1, 1–8). Significantly, these ITC results also reveal previously uncharacterized DNA binding properties of the CXXC domain. First and most importantly, the Tet3 CXXC domain strongly binds to both non-CpG (Table S1, 5–7) and CpG DNA oligos (Table S1, 1–4 and 9) with a slight preference for CpG DNA oligos. Second, the Tet3 CXXC domain strongly binds to CmCGG DNA (Figure 4F and Table S1, 10). These binding properties are in stark contrast to the well-characterized DNA binding property of the CFP1 and MLL CXXC domains, which absolutely require unmethylated CpG dinucleotides (Allen et al., 2006; Xu et al., 2011a). In contrast to the strong binding to CmCGG, CTAG and ACGT DNA oligos, the Tet3 CXXC domain does not bind to mCmCGG, mCTAG, or AmCGT DNA oligos (Figure 4G and Table S1, 11–13). These data suggest that although strict CpG content is not required for the Tet3 CXXC domain to interact with DNA, an unmodified cytosine is

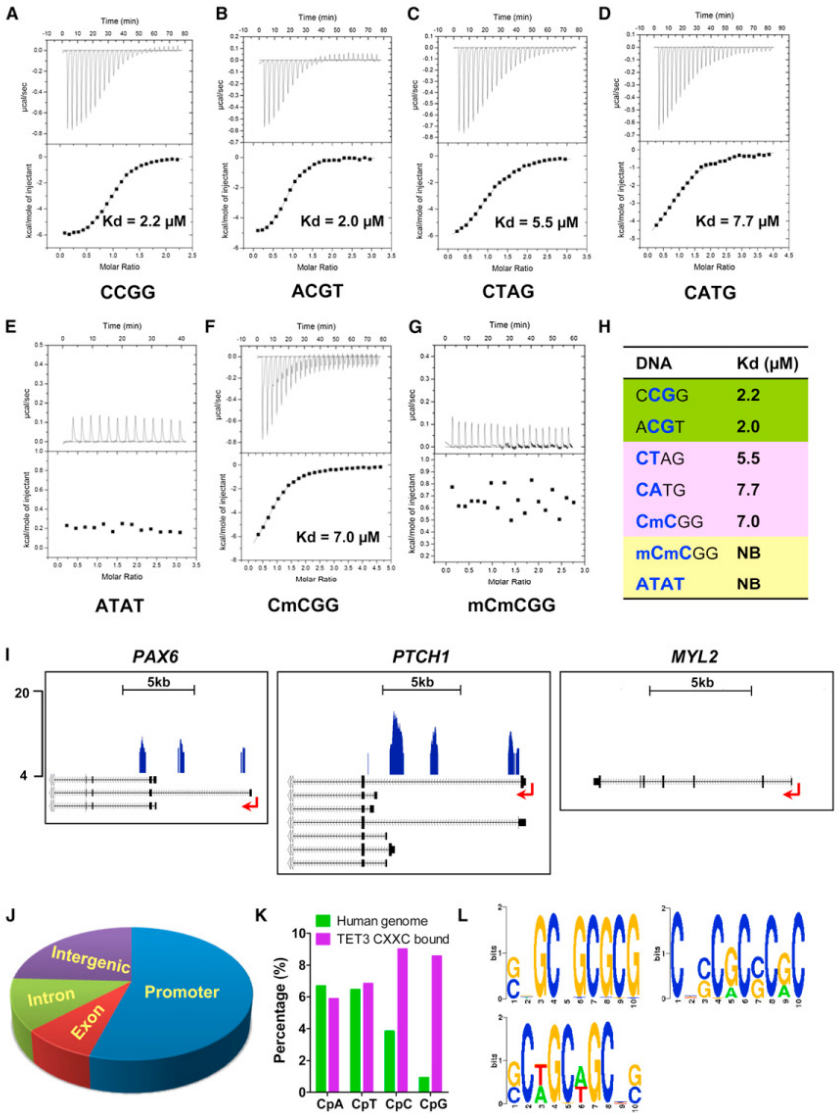


Figure 4. The Unique DNA Binding Properties of the Tet3 CXXC Domain

(A–H) Binding affinities of the Tet3 CXXC domain to various DNA oligos by ITC assays. The sequence of the central four nucleotides of each double-stranded DNA probe is shown under the corresponding panel. Detailed sequence information for all DNA oligos used in this study is listed in Table S1. NB: no binding.

(I) Representative TET3 CXXC GST pull-down sequencing results. Arrow denotes promoter orientation.

(J) Genomic distribution of the TET3 CXXC-bound loci. Promoter is defined as −2 kb to +2 kb relative to transcription start site (TSS).

(K) The percentage of CpA, CpT, CpC, and CpG in human genome and the TET3 CXXC-bound loci.

(L) DNA motifs that are enriched in the TET3 CXXC-bound loci.

See also Figure S4 and Table S1.

essential for the Tet3 CXXC binding to DNA. Using the same ITC approach, we also demonstrate that the human TET3 CXXC domain has similar DNA binding properties to the *Xenopus* Tet3 CXXC domain (Table S1, 17–22). Thus, the CXXC domain of the Tet3 family has unique newly identified DNA binding characteristics conserved among vertebrates.

Given the similar binding properties between human and *Xenopus* Tet3 CXXC domains, we next determined the DNA binding specificity of the TET3 CXXC domain across the whole genome of HEK293T cells. The specific TET3 CXXC-bound genomic DNA fragments were enriched by GST pull-downs and analyzed by deep DNA sequencing, as we previously described (Xu et al., 2011b). The TET3 CXXC domain selectively binds to restricted genomic regions (Figure S4B). Importantly, consistent with Tet3 occupancy in *Xenopus* embryos (Figure 2C), the TET3 CXXC domain strongly and selectively binds to the promoters of *PAX6*, *PTCH1*, *NGN2*, *TUBB2B*, and *SHH*, but not the promoter of *MYL2* (Figures 4I and S4C). Furthermore, bioinformatics analyses identify 17,953 TET3 CXXC-bound peaks, more than half of which are located at gene promoters (Figure 4J). Interestingly, we also observe a significant enrichment of CpG (and a less degree of CpC) dinucleotides in TET3 CXXC-bound regions (Figure 4K). Further de novo motif discovery analyses identify three C-rich sequences among the top 15 ranking motifs within TET3 CXXC-bound regions (Figure 4L), whose consensus sequences are SSGCSGCGCG ($p = 1 \times 10^{-30}$), CSSCGCSCRC ($p = 3.4711 \times 10^{-26}$) and SCWGCWGCBS ($p = 4.6563 \times 10^{-25}$), respectively. Indeed, we validate that these motifs are present at the promoters of several TET3 target genes (Figure S4D). Together with the ITC binding data, these genome-wide analyses suggest that the TET3 CXXC domain is able to bind to the unmodified C followed by A, T, C, or G with a slight preference for CpG dinucleotides. These data also indicate that the specific DNA binding activity of the Tet3 CXXC domain may contribute to Tet3 targeting, thereby serving as another important mechanism for Tet3-mediated gene transcriptional regulation.

Table 1. Data Collection and Refinement Statistics

	xtTet3 CXXC-GCCAACGTTGGC	xtTet3 CXXC-GCCACmCGGTGGC
PDB code	4HP3	4HP1
Space group	C2	C2
Cell Dimensions		
a, b, c (Å)	69.9, 39.5, 54.1	71.0, 39.5, 57.6
α, β, γ (°)	90, 99.9, 90	90, 91.0, 90
Resolution range (Å)	34.42-2.05 (2.16-2.05)	35.51-2.25 (2.37-2.25)
Number of unique HKLs	9281 (1345)	7738 (1110)
Completeness (%)	99.8 (99.8)	99.9 (100.0)
Friedel redundancy	3.7 (3.7)	4.1 (4.2)
Rsym (%)	5.3 (66.9)	4.6 (48.7)
$<I/\sigma(I)>$	12.6 (2.1)	14.7 (2.8)
Resolution limits (Å)	30.00-2.05	35.00-2.25
Number of unique HKLs work/free	8829/452	7385/351
R_{work}/R_{free} (%)	21.6/24.3	22.0/25.0
Number of atoms/< B > ($Å^2$)	898/54.8	886/73.9
DNA	486/58.2	487/78.8
Protein	384/51.7	387/68.5
Zn^{2+}	2/33.2	2/46.8
RMSD bonds (Å)/angles (°)	0.013/1.4	0.012/1.4
Ramachandran plot favored residues, no outliers (Lovell et al., 2003)	47 of 49	48 of 49

RMSD is an abbreviation for root mean squared deviation. Average B-factors calculated with MOLEMAN (G.J. Kleywegt, Uppsala University). The highest resolution shell is shown in parentheses.

Crystal Structures of the Tet3 CXXC Domain in Complex with DNA Oligos

To gain further mechanistic insight into the unique DNA binding properties of the Tet3 CXXC domain, we next determined the crystal structures of the *Xenopus* Tet3 CXXC domain (aa 58–111) in complex with 5mC-containing DNA (CmCGG) and CpG-containing DNA (ACGT), respectively (Table 1). Like other CXXC domains, the Tet3 CXXC domain contains eight conserved cysteine residues coordinating two zinc ions. These two zinc ions play a structural role by holding the mainly unstructured CXXC domain together and forming a crescent-shaped architecture to bind DNA (Figures 5A and 5C). The DNA-binding surface is predominantly positively charged, and wedged into the major groove of DNA to extensively interact with DNA bases by means of hydrogen bonds and electrostatic interactions (Figures 5A–5D).

Surprisingly, structural comparison of the Tet3 CXXC domain in complex with CmCGG or ACGT DNA reveals that the Tet3 CXXC domain binds to the 12-mer target DNAs with one nucleotide shift. In the crystal structure of the Tet3 CXXC domain in complex with the unmethylated CpG DNA (ACGT), the Tet3 CXXC domain binds to the DNA centering on the cytosine in the CpG dinucleotide (underlined). Methylation of the cytosine in the CpG dinucleotides (CmCGG) shifts the binding to the cytosine (underlined) preceding the mCpG dinucleotides. Therefore, consistent with the ITC binding results, the Tet3 CXXC domain binds to an unmodified cytosine, which is not restricted by CpG content, distinct from the CpG-dependent binding of the MLL, CFP1, and DNMT1 CXXC domains (Allen et al., 2006; Pradhan et al., 2008; Xu et al., 2011a).

In both Tet3 CXXC-ACGT and Tet3 CXXC-CmCGG complex structures, the main chain carbonyl oxygen of residue His90 forms a hydrogen bond with the target cytosine, whereas the side chain of His90 forms another hydrogen bond with the complementary guanine (Figures 5E, 5G, S5A and S5C). In the Tet3 CXXC-ACGT complex structure, residues Gln91 and Ser89 form hydrogen bonds with the guanine-cytosine base pair following the target cytosine-guanine base pair (Figures 5F and S5B). In contrast, in the Tet3 CXXC-CmCGG complex structure, methylation of C6 introduces steric clashes with the side chain of Gln91 and causes the side chain of Gln91 to become partially disordered. As a result, Ser89 flips peptide plane and

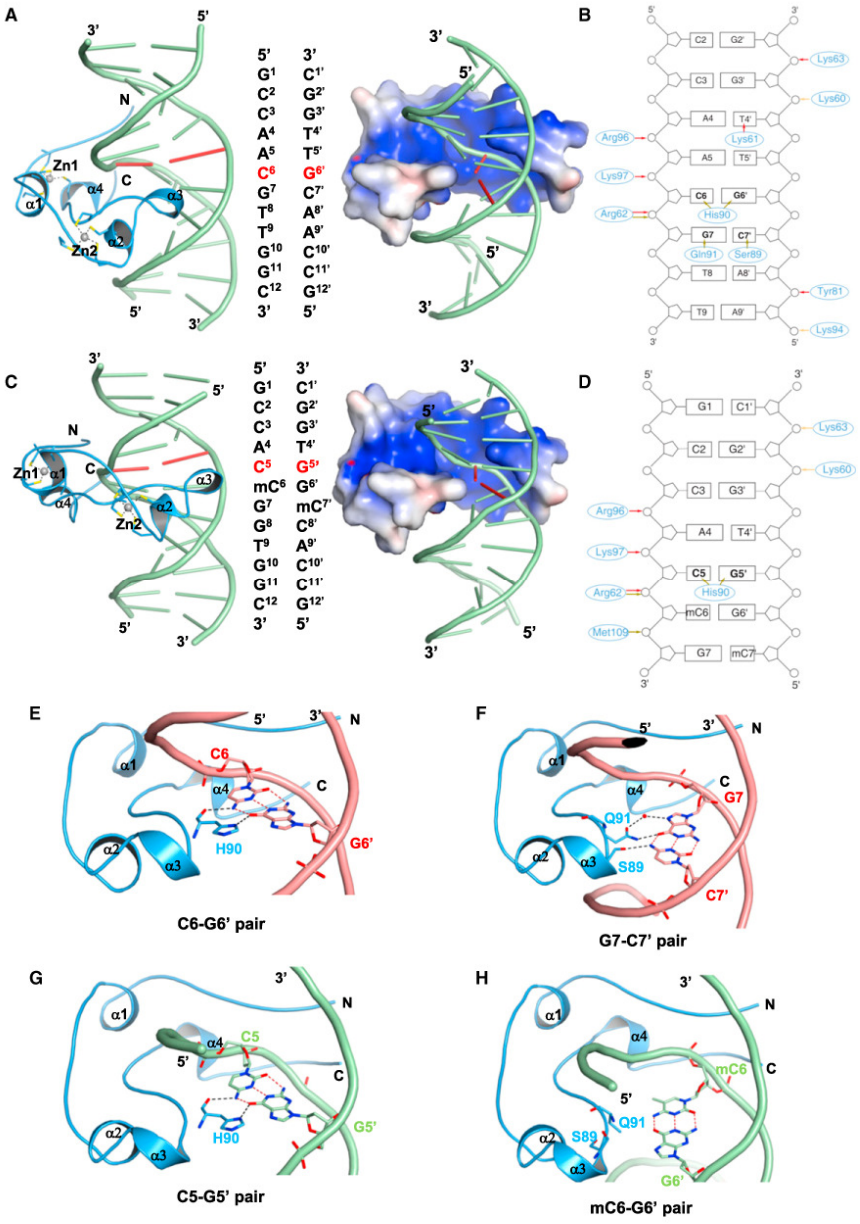


Figure 5. The Tet3 CXXC Domain Specifically Recognizes Cytosine through a Conserved Residue His90

(A and C) Crystal structures of the Tet3 CXXC domain in complex with ACGT DNA (5′-GCCA**ACGT**TGGC-3′) (A) or CmCGG DNA (5′-GCCA**CmCGG**TGGC-3′) (C) in cartoon (left) and electrostatic representations (right), respectively. The double-stranded DNA sequence is shown in the middle of each corresponding panel.

loses its hydrogen bond interaction with DNA (Figures 5H and S5D). These data provide structural explanations for the binding to CmCGG DNA and the slight binding preference for CpG compared to CpH DNAs (H = A, T, C, or 5mC). Substituting the target cytosine with any other nucleotide or 5mC would introduce steric clashes with His90 (Figure S5E), consistent with the undetectable binding of the Tet3 CXXC domain to ATAT DNA (Figure 4E). To determine the role of residue His90 of Tet3, which is conserved among all Tet CXXC domains (Figure S4A), we generated the Tet3 CXXC H90A mutant and tested its DNA binding activities by ITC assay. As expected, the Tet3 CXXC H90A mutant loses its DNA binding abilities (Table S1, 14–15), supporting an essential role of His90 in DNA binding.

To further understand the novel DNA binding property of the Tet3 CXXC domain, we superimposed the complex structures of the Tet3 and CFP1 CXXC domains (Xu et al., 2011a) (Figure S6A). Although the Tet3 CXXC domain adopts a similar fold to that of the CFP1 CXXC domain, the loop region preceding His90 of Tet3 is dramatically different from that of CFP1 (Figure S6A). In the CFP1 CXXC domain, Asp189 forms three hydrogen bonds with the backbone of Lys198, Ile199, and Arg200, making the loop very rigid and only allowing CpG binding (Xu et al., 2011a). Notably, this Asp residue is highly conserved in the CXXC domains of CFP1, MLL, DNMT1, and KDM2A (Figure S4A). However, the corresponding loop between the α3 and α4 helices of the Tet3 CXXC domain is much shorter and more flexible due to lack of hydrogen bonds seen in the rigid loop of the CFP1 CXXC domain (Figure S6A). Therefore, the shorter and less rigid loop of the Tet3 CXXC domain can accommodate other nucleotides besides G, such as T, C, A, or 5mC, following the target C, which confers cytosine-specific instead of CpG-specific binding property. In addition, we also observe that the Tet3 CXXC domain binds to TCGA and mCCGG DNA oligos (Table S1, 2 and 9) with slightly weaker binding affinities compared to other CpG DNA oligos (Table S1, 1, 3 and 4). Although the methyl group of thymine or 5mC does not cause any steric clash with the Tet3 CXXC domain, the hydrophobic methyl group points to the solvent and no residues in the Tet3 CXXC domain can recognize and stabilize it, which is not energetically favorable (Figures S6B–S6E). Taken together, the Tet3 CXXC domain binds to cytosine-containing DNA with a slight preference for a G following the target cytosine and modestly disfavoring a T or mC preceding the target cytosine.

Both 5mC Hydroxylase Activity and CXXC Domain-Mediated Specific DNA Binding Are Required for Tet3 Function in Target Gene Regulation and Embryonic Development

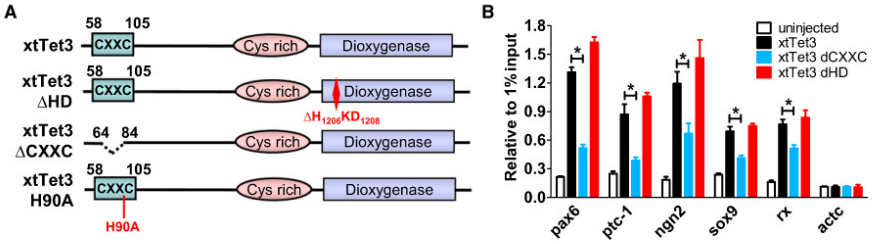
Having defined the regulation of 5mC/5hmC status at Tet3 target gene promoters by the 5mC hydroxylase activity and the unique DNA binding properties of the Tet3 CXXC domain, we next determined the role of these two functional domains in Tet3 function. We generated wild-type (*xtTet3*), CXXC-deleted (*xtTet3ΔCXXC*), His90-to-Ala (*xtTet3H90A*), and iron binding site-disrupted catalytically inactive (*xtTet3ΔHD*) *xtTet3* expression constructs (Figure 6A). We confirm that both the wild-type and mutant xtTet3 proteins are properly expressed and localized to the nucleus (Figures S7A and S7B) and that the 5mC hydroxylase activity is retained in xtTet3ΔCXXC and xtTet3H90A but not in xtTet3ΔHD (Figure S7B). We also performed ChIP-qPCR assays to compare the occupancy of wild-type and mutant Tet3 proteins at target gene promoters. Both the CXXC domain deletion and H90A point mutation abolish the occupancy of mutant proteins at target gene promoters (Figures 6B and S7C), suggesting an essential role of the CXXC domain-mediated DNA binding in Tet3 targeting to specific genes. Importantly, although both xtTet3 and xtTet3ΔHD exhibit similar occupancy, coinjection of *xtTet3*, but not *xtTet3ΔHD*, significantly rescues the decreased 5hmC levels caused by Tet3 depletion at target gene promoters including the *rx* gene promoter (Figure S7D), suggesting that the enzymatic activity of Tet3 is primarily responsible for the dynamic regulation of 5hmC at its target genes.

We next used these constructs to determine which domain(s) are critical for Tet3 functions in vivo. We first employed a functional domain rescue approach to interrogate the role of Tet3 catalytic domain using *pax*6 gene expression as a readout. The expression of *pax6* is completely rescued in 79% of *xtTet3* coinjected embryos, whereas the complete rescue effect of *xtTet3 ΔHD* coinjection only reaches 37%, significantly lower than that of *xtTet3* coinjection ($p < 0.01$) (Figure 6C). These results are corroborated by phenotypic rescues. Only 25% of embryos are completely rescued by *xtTet3 ΔHD* coinjection, again significantly lower than that of *xtTet3* coinjection, in which 62% are completely rescued ($p < 0.01$) (Figure 6D). Thus, these data suggest that the 5mC hydroxylase activity of Tet3 not only controls the dynamics of 5mC/5hmC at target gene promoters but is also required for target gene regulation and biological function in early embryonic development. Next, we employed the same strategy to examine the importance of the CXXC domain/His90-mediated specific DNA binding in Tet3 function. Strikingly, *xtTet3 H90A* or *xtTet3 ΔCXXC* coinjection shows no rescue effects on the inhibited *pax6* expression or developmental phenotypes caused by Tet3 depletion (Figure 6C and 6D), highlighting an essential role of the CXXC domain-mediated DNA binding in Tet3 function in target gene regulation and embryonic development.

DISCUSSION

In this study, we report that *Xenopus* Tet3 is a new class of transcription regulator playing an essential role in early eye and neural development by directly regulating a set of key genes critical for these developmental processes. We also uncover that

(B and D) Detailed interactions between the Tet3 CXXC domain and ACGT DNA (B) or between the Tet3 CXXC domain and CmCGG DNA (D). Red arrow, salt bridge interaction; olive arrow, hydrogen bond interaction; yellow arrow, electrostatic interaction.
(E and F) Detailed interactions between the Tet3 CXXC domain (blue cartoon) and the target CG pair (E) or the following GC pair (F) in the ACGT DNA (red cartoon).
(G and H) Detailed interactions between the Tet3 CXXC domain (blue cartoon) and the target CG pair (G) or the following mCG pairs (H) in the CmCGG DNA (green cartoon).
See also Figures S5 and S6.



C

***pax6* gene expression rescue**

Group	Suppressed	Partially suppressed	Normal	Total	Note
1. Control MO	0% (n=0)	0% (n=0)	100% (n=31)	53	
2. *xlTet3* MOs	96% (n=53)	0% (n=0)	4% (n=2)	55	
3. *xlTet3* MOs/*xtTet3*	7% (n=4)	14% (n=8)	79% (n=44)	56	**
4. *xlTet3* MOs/*xtTet3ΔCXXC*	92% (n=50)	4% (n=2)	4% (n=2)	54	#
5. *xlTet3* MOs/*xtTet3ΔHD*	40% (n=21)	23% (n=12)	37% (n=19)	52	**, #
6. *xlTet3* MOs/*xtTet3H90A*	93% (n=52)	2% (n=1)	5% (n=3)	52	#

D

Phenotype rescue

Group	Complete defect	Partial defect	Normal	Total	Note
1. uninjected	0% (n=0)	0% (n=0)	100% (n=183)	183	
2. Control MO	0% (n=0)	0% (n=0)	100% (n=192)	192	
3. *xlTet3* MOs	85% (n=171)	11% (n=23)	4% (n=7)	201	
4. *xlTet3* MOs/*xtTet3*	20% (n=38)	18% (n=36)	62% (n=121)	195	**
5. *xlTet3* MOs/*xtTet3ΔCXXC*	80% (n=148)	12% (n=22)	8% (n=15)	185	#
6. *xlTet3* MOs/*xtTet3ΔHD*	39% (n=73)	36% (n=68)	25% (n=48)	189	**, #
7. *xlTet3* MOs/*xtTet3H90A*	80% (n=151)	14% (n=27)	6% (n=12)	190	#

E

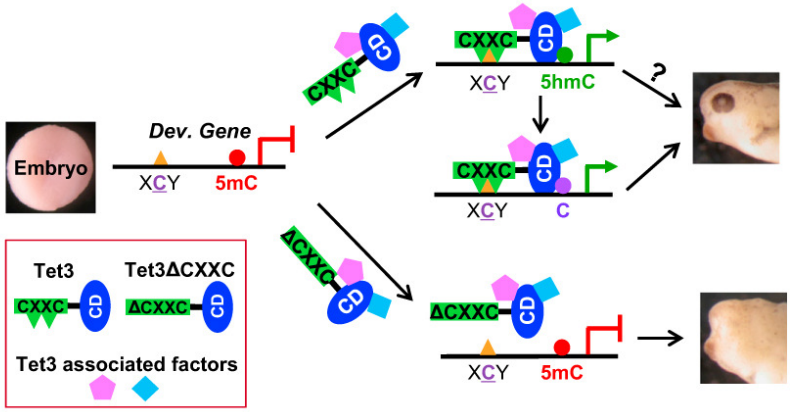


Figure 6. Both 5mC Hydroxylase Activity and the CXXC Domain Are Important for Tet3 Function

(A) Schematic representation of xtTet3 mutants.

(B) The CXXC domain deletion disrupts Tet3 occupancy at specific gene promoters by ChIP-qPCR assay. Data are presented as mean ± SEM (n = 3). *p < 0.05.

(C) Summarized results of five independent *pax6* expression rescue experiments. "Suppressed" means significantly suppressed *pax6* expression in posterior and anterior neural plates; "Partially suppressed" means *pax6* expression is detected but not intact in posterior and anterior neural plates; "Normal" means intact *pax6* expression in posterior and anterior neural plates. **p < 0.01 compared to *xlTet3* MOs, # p < 0.01 compared to *xlTet3* MOs/*xtTet3*.

the dynamic regulation of 5mC/5hmC status at target gene promoters is an important mechanism underlying Tet3-mediated target gene regulation during embryonic development. The Tet3 CXXC domain-mediated specific DNA binding is essential for targeting Tet3 to its target genes, thus, providing another layer of regulation on the transcription of Tet3 target genes. Our findings support a molecular model of Tet3 action in target gene regulation that involves Tet3 binding to unmethylated cytosines (with a slight preference for CpG content) at target gene promoters through its CXXC domain and its intrinsic 5mC hydroxylase activity converting adjacent 5mC to 5hmC, which is an intermediate for further DNA demethylation (Figure 6E). Recruitment to a specific gene promoter and subsequent conversion of 5mC to 5hmC cooperatively activate the expression of Tet3 target genes, including those identified key developmental genes, to ensure normal and precisely regulated embryogenesis. Inactivation of either one of these functional domains therefore will have adverse impact on Tet3 function. Notably, our data also show that the catalytically inactive Tet3 mutant retains partial rescue effects in vivo, suggesting potential enzymatic activity-independent mechanisms, such as recruiting or interacting with other transcription factors, may also contribute to the full function of Tet3 in gene regulation (Figure 6E). Indeed, TET3 forms a stable complex with several critical transcription factors and histone modifiers (data not shown). It has also been shown that an enzymatic activity-independent mechanism is involved in gene regulation by Tet1 in mouse ES cells (Williams et al., 2011). Moreover, it has been suggested that 5hmC is a stable epigenetic mark that participates in regulating gene expression through unidentified mechanisms such as recruiting unidentified 5hmC specific "readers" (Branco et al., 2012; Matarese et al., 2011). Thus, this model does not exclude the possibility that Tet3 and 5hmC may participate in target gene regulation through this yet to be identified regulatory circuit.

Unlike well-characterized DNA methylation (5mC) and DNMTs, the mechanism by which 5hmC and the enzymatic activity of TET proteins contribute to gene regulation has been elusive. Our findings here offer a biological model demonstrating that the dynamic regulation of 5mC/5hmC by the Tet family of 5mC hydroxylases has an important role in gene regulation during early eye and neural development. We show that Tet3 is an active 5mC hydroxylase highly expressed in the region of the developing brain, eye, and spinal cord in *Xenopus* embryos. As exemplified by *rx* and *pax6*, our data clearly show that Tet3 target genes have significant alterations in 5hmC status at their promoters after Tet3 depletion. Importantly, 5hmC level alterations and dysregulated gene expression can be completely rescued by wild-type Tet3. Yet the rescue effect by the catalytically inactive Tet3 mutant is significantly impaired, highlighting the importance of 5mC hydroxylase activity in Tet3 function. Furthermore, consistent with the eye developmental phenotypes in *Xenopus*, we find that homozygous deletion of the Tet3 catalytic domain in pure B6 genetic background mice (Gu et al., 2011) results in eye-related phenotypes including the eyelid open at birth (EOB) phenotype (data not shown), suggesting an evolutionally conserved function in eye development from vertebrate to mammals and the critical role of the 5mC hydroxylase activity in Tet3 function. Taken together, our study clearly indicates that precise regulation of the dynamics of DNA modification status at specific gene loci by Tet3-mediated conversion of 5mC to 5hmC is an important and conserved epigenetic mechanism for target gene regulation.

How epigenetic enzymes are targeted and/or confined to their functional sites is a fundamental question in understanding the mechanism underlying epigenetic transcription regulation. In this study, our biochemical characterizations reveal that the Tet3 CXXC domain binds to DNA in a cytosine-dependent manner with a slight preference for CpG dinucleotides, distinct from the CpG-dependent binding of other well-characterized CXXC domains. We also notice that the DNA binding property of the Tet3 CXXC domain is different from that of the TET1 CXXC domain that binds to unmodified C or 5mC- or 5hmC-modified CpG-rich DNA (Xu et al., 2011b), suggesting that Tet1 and Tet3 may have different functions in gene regulation besides their distinct expression patterns during early embryonic development (Tan and Shi, 2012). Moreover, our genome-wide mapping and de novo motif analyses of the TET3 CXXC domain binding sites strongly suggest that the TET3 CXXC domain selectively binds to several consensus sequences, therefore offering a mechanism for targeting Tet3 to its target genes. Finally, the crystal structure analysis of the Tet3 CXXC domain in complex with DNA provides the following structural mechanisms underlying its novel DNA binding ability and specificity. (1) It reveals that the major binding force of the Tet3 CXXC domain to CmCGG DNA is through the specific interaction between residue His90 and the unmethylated cytosine, which explains why the Tet3 CXXC domain can still bind to partially methylated DNA, as long as there is an unmethylated cytosine available in the target sequence; (2) it explains the slight binding preference of the Tet3 CXXC domain for CpG compared to CpH (H = A, T, C, or 5mC) and the modest disfavor for a T or 5mC preceding the target C, suggesting that sequences flanking the target C contribute to the overall binding affinity and specificity of the Tet3 CXXC domain.

The importance of this CXXC domain-mediated specific DNA binding activity of Tet3 is further demonstrated in our functional studies using the Tet3 mutants with the CXXC domain deletion (ΔCXXC) or H90A point mutation. These mutants lose the specific association with Tet3 target gene promoters and functional rescue abilities in vivo. Taken together, our study reports a novel DNA binding property and functionality of the CXXC

(D) Summarized results of five independent phenotypic rescue experiments. "Partial defect" means mild abnormal head structure, small eyes or one eye; "Complete defect" means abnormal head structure and no eye. **p < 0.01 compared to *xlTet3* MOs, # p < 0.01 compared to *xlTet3* MOs/*xtTet3*.

(E) A model of Tet3 action in gene transcription regulation. The Tet3 CXXC domain specifically binds to unmodified cytosine (underlined)-containing sequence motifs with a slight preference for G at "Y" position and a mild disfavor for T or 5mC at "X" position, targeting Tet3 to the promoter of target developmental genes. Then, the 5mC hydroxylase activity of Tet3 converts adjacent 5mC to 5hmC, an intermediate for further DNA demethylation, thus activates the gene expression. CD: catalytic domain. Please refer to the related text for more details.

See also Figure S7.

domain that is essential for Tet3 function both in gene regulation and embryonic development. It not only provides a key mechanistic layer of Tet3-mediated target gene regulation but also significantly advances our current understanding of the molecular and biological function of the Tet family CXXC domains. Noteworthy, even though the CXXC domain is essential for targeting Tet3 to specific genomic regions, we do not exclude the involvement of other potential cellular mechanisms for targeting or recruiting Tet3 to its functional sites. We favor the hypothesis that Tet3 is likely in complex with many other cellular factors including sequence-specific DNA binding transcription factors and cofactors, to execute its molecular, cellular, and biological functions. Therefore, the associated transcription factors may in part coordinate with the Tet3 CXXC domain and enzymatic activity and contribute to the overall mechanism of action of Tet3 in gene regulation and embryonic development.

EXPERIMENTAL PROCEDURES

Embryo Manipulation and Microinjection

X. laevis eggs were artificially fertilized with testis homogenate and cultivated in 0.1× MMR as previously described (Sakano et al., 2010). Capped synthetic mRNAs were generated by in vitro transcription with sp6 polymerase. Embryos were transferred to 3% Ficoll 400 in 0.1× MMR and injected embryos were cultured in 0.1× MMR until the desired stage. For phenotype experiments, 80 ng of *xlTet3* MOs (40ng of each *xlTet3* MO) or control MO were injected into two dorsal blastomeres of 4-cell stage embryos. For phenotype rescue experiments, 1 ng of mRNA was coinjected with 80 ng of *xlTet3* MOs into two dorsal blastomeres of 4-cell stage embryos. For *pax6*, *rx*, *six3*, *sox2*, *otx2*, *sox9*, *snail*, *tubb2b*, and *ngn2 in situ* hybridization experiments, 160 ng of *xlTet3* MOs or control MO were injected into one dorsal blastomere of each 4-cell stage embryo. Because the expression of *shh* and *ptc-1* is on the midline, for *shh* and *ptc-1 in situ* hybridization experiments, two dorsal blastomeres of each 4-cell stage embryo were injected with 160 ng of *xlTet3* MOs or control MO. In all injection studies, 200 pg of *nucβ-gal* RNA were coinjected as the injection tracer.

Crystallization

Each pair of single-stranded DNAs was mixed with a molar ratio of 1:1 and annealed to form double-stranded DNAs. Before cocrystallization, purified xtTet3 CXXC protein was mixed with different DNAs in a molar ratio of 1:1.2. Crystals of xtTet3 CXXC in complex with ACGT DNA (GCCAACGTTGGC) were obtained via sitting drop vapor diffusion; 1.0 μl of complex was mixed with 1.0 μl of well solution containing 0.1 M HEPES (pH 7.5), 0.2 M NaCl, 30% PEG 1500, against 800 μl of reservoir buffer at 18°C. Crystals grow to a mountable size in 3 days. Crystals of xtTet3 CXXC in complex with CmCGG DNA (GCCACmCGGTGGC) were obtained in a similar way in the buffer containing 0.1 M HEPES (pH 7.5), 0.2 M NaCl, 30% PEG 1500, 5% MPD. Both crystals were flash-frozen in liquid nitrogen directly without cryoprotectant.

ACCESSION NUMBERS

The sequences of Tet3 genes have been deposited into GeneBank under the accession numbers of HQ220209 (human TET3), HQ423151 (mouse Tet3), HQ220207, and HQ220208 (*X. laevis* Tet3), respectively. The structures of Tet3 CXXC-DNA complexes have been deposited in the Protein Data Bank under accession number 4HP1 and 4HP3. The TET3 CXXC GST pull-down sequencing data have been deposited in GEO database under the accession number GSE41551.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2012.11.014.

ACKNOWLEDGMENTS

We thank R. Harland for helpful discussions, A Khorasani for critical reading of the manuscript and Q. Xu and L. Jiang for statistical analyses. We also thank R. Moon, W. Harris, J. Yang, X. Wang, M. Vetter and J.-P. Saint-Jeannet for valuable discussions and technical help. This work was supported by NIH grants GM078458 to Y.G.S., GM087641 to Y.K., and GM074241 to X.H., Brigham and Women's Hospital Biomedical Research Institute fund to sustain research excellence to Y.G.S., and Chinese Ministry of Education Project 985 to Biomedical Core Facility, Fudan University. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. J.G.A. was funded partially by CNPq (Brazil). Y.G.S. is a Pew Scholar. X.H. acknowledges support by Boston Children's Hospital Intellectual and Developmental Disabilities Research Center (P30 HD-18655). The Structural Genomics Consortium is a registered charity (No. 1097737) that receives funds from Canadian Institutes for Health Research, Canadian Foundation for Innovation, Genome Canada through the Ontario Genomics Institute, GlaxoSmithKline, Eli Lilly, Pfizer, Novartis Research Foundation, Life Technologies, Ontario Innovation Trust, Ontario Ministry for Research and Innovation and Wellcome Trust (JM). Argonne National Laboratory is operated by UChicago Argonne, LLC, for the U.S. Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357.

Received: May 17, 2012
Revised: September 19, 2012
Accepted: November 9, 2012
Published: December 6, 2012

REFERENCES

Allen, M.D., Grummitt, C.G., Hilcenko, C., Min, S.Y., Tonkin, L.M., Johnson, C.M., Freund, S.M., Bycroft, M., and Warren, A.J. (2006). Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. EMBO J. *25*, 4503–4512.

Bestor, T.H., and Coxon, A. (1993). Cytosine methylation: the pros and cons of DNA methylation. Curr. Biol. *3*, 384–386.

Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. Nature *321*, 209–213.

Branco, M.R., Ficz, G., and Reik, W. (2012). Uncovering the role of 5-hydroxymethylcytosine in the epigenome. Nat. Rev. Genet. *13*, 7–13.

Delhommeau, F., Dupont, S., Della Valle, V., James, C., Trannoy, S., Massé, A., Kosmider, O., Le Couedic, J.P., Robert, F., Alberdi, A., et al. (2009). Mutation in TET2 in myeloid cancers. N. Engl. J. Med. *360*, 2289–2301.

Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature *473*, 398–402.

Globisch, D., Münzel, M., Müller, M., Michalakis, S., Wagner, M., Koch, S., Brückl, T., Biel, M., and Carell, T. (2010). Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS ONE *5*, e15367.

Gu, T.-P., Guo, F., Yang, H., Wu, H.-P., Xu, G.-F., Liu, W., Xie, Z.-G., Shi, L., He, X., Jin, S.G., et al. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature *477*, 606–610.

He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303–1307.

Hsu, C.-H., Peng, K.-L., Kang, M.-L., Chen, Y.-R., Yang, Y.-C., Tsai, C.-H., Chu, C.-S., Jeng, Y.-M., Chen, Y.-T., Lin, F.-M., et al. (2012). TET1 Suppresses Cancer Invasion by Activating the Tissue Inhibitors of Metalloproteinases. Cell Rep. *2*, 568–579.

Iqbal, K., Jin, S.-G., Pfeifer, G.P., and Szabó, P.E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proc. Natl. Acad. Sci. USA *108*, 3642–3647.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129–1133.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300–1303.

Koh, K.P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., Laiho, A., Tahiliani, M., Sommer, C.A., Mostoslavsky, G., et al. (2011). Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell Stem Cell *8*, 200–213.

Langemeijer, S.M., Kuiper, R.P., Berends, M., Knops, R., Aslanyan, M.G., Massop, M., Stevens-Linders, E., van Hoogen, P., van Kessel, A.G., Raymakers, R.A., et al. (2009). Acquired mutations in TET2 are common in myelodysplastic syndromes. Nat. Genet. *41*, 838–842.

Lian, C.G., Xu, Y., Ceol, C., Wu, F., Larson, A., Dresser, K., Xu, W., Tan, L., Hu, Y., Zhan, Q., et al. (2012). Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. Cell *150*, 1135–1146.

Matarese, F., Carrillo-de Santa Pau, E., and Stunnenberg, H.G. (2011). 5-Hydroxymethylcytosine: a new kid on the epigenetic block? Mol. Syst. Biol. *7*, 562.

Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M.E., Vasanthakumar, A., Patel, J., Zhao, X., et al. (2011). Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell *20*, 11–24.

Pradhan, M., Estève, P.-O., Chin, H.G., Samaranayke, M., Kim, G.-D., and Pradhan, S. (2008). CXXC domain of human DNMT1 is essential for enzymatic activity. Biochemistry *47*, 10000–10009.

Quivoron, C., Couronné, L., Della Valle, V., Lopez, C.K., Plo, I., Wagner-Ballon, O., Do Cruzeiro, M., Delhommeau, F., Arnulf, B., Stern, M.-H., et al. (2011). TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. Cancer Cell *20*, 25–38.

Reik, W., Dean, W., and Walter, J. (2001). Epigenetic reprogramming in mammalian development. Science *293*, 1089–1093.

Sakano, D., Kato, A., Parikh, N., McKnight, K., Terry, D., Stefanovic, B., and Kato, Y. (2010). BCL6 canalizes Notch-dependent transcription, excluding Mastermind-like1 from selected target genes during left-right patterning. Dev. Cell *18*, 450–462.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930–935.

Tan, L., and Shi, Y.G. (2012). Tet family proteins and 5-hydroxymethylcytosine in development and disease. Development *139*, 1895–1902.

Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A.C., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature *473*, 343–348.

Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C.J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., and Walter, J. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat. Commun. *2*, 241.

Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature *473*, 389–393.

Xu, C., Bian, C., Lam, R., Dong, A., and Min, J. (2011a). The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain. Nat. Commun. *2*, 227.

Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A.J., Zheng, L., Zhang, H., Huang, S., et al. (2011b). Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. Mol. Cell *42*, 451–464.

Yu, M., Hon, G.C., Szulwach, K.E., Song, C.-X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., et al. (2012). Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell *149*, 1368–1380.